PB85-137239

Summary Report of the Workshop on Structure Activity Concepts in Environmental Sciences

(U.S.) Environmental Research Lab.-Duluth, MN

Dec 84

U.S. Department of Commerce
National Technical Information Service
NTIS

PB85-137339

EPA-600/D-84-284
December 1984

SUMMARY REPORT OF THE WORKSHOP ON STRUCTURE ACTIVITY CONCEPTS IN ENVIRONMENTAL SCIENCES

Edited by

Marion W. Meader

EPA Project Officer
G. Veith

ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
DULUTH, MN 55804

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO.
EPA-600/D-84-284

2.

3. RECIPIENT'S ACCESSION NO.
PB85 137339

4. TITLE AND SUBTITLE
Summary Report of the Workshop on Structure Activity Concepts in Environmental Sciences

5. REPORT DATE
December 1984

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
R. Haque and G. Veith* (Chairmen)

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
*U.S. Environmental Protection Agency
Environmental Research Laboratory-Duluth
6201 Congdon Boulevard
Duluth, Minnesota 55804

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.

12. SPONSORING AGENCY NAME AND ADDRESS
Toxics and Pesticides Division
Office of Environmental Processes and Effects Research
Office of Research and Development
U.S. Environmental Protection Agency

13. TYPE OF REPORT AND PERIOD COVERED

14. SPONSORING AGENCY CODE
EPA-600/03

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The theoretical and practical considerations in the application of structure-activity relationships (SAR) to chemical hazard assessment have been the subject of much research and debate. In order to focus this research on producing the most cost-effective methods of chemical evaluation, EPA sponsored a workshop on SAR in Environmental Sciences in San Antonio, Texas. Experts from industry, academia and government were invited to examine the state-of-the-art of current SAR methods, the limitations of SAR for EPAs mission, and develop research priorities for SAR over the next two years. This report is a summary of the workshop findings.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT
Release to public

19. SECURITY CLASS (This Report)
Unclassified

20. SECURITY CLASS (This page)
Unclassified

21. NO. OF PAGES
308

22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iii
EXECUTIVE SUMMARY	1
CONCLUSIONS AND RECOMMENDATIONS	25
REPORTS OF WORKING GROUPS	43
Group I - Available Methods for Structure Activity: Requirements and Limitations	
Group II - Evaluation of Molecular Descriptors	
Group III - Structure Activity Relationships for Environmental Fate Assessment	
Group IV - Predicting Environmental Effects	
Group V - Structure-Activity Applications in Regulation	
APPENDIX - Program	A-1
List of Attendees	A-2

LIST OF FIGURES

FIGURE NUMBER		
1	Structure to Hazard and Decisions	19

EXECUTIVE SUMMARY

Of the several hundred billion pounds of synthetic chemicals manufactured annually in this country, a significant amount enters the air, water, and soil, often at very low concentrations. Some of these chemicals persist, accumulate, and bioconcentrate, thereby affecting plant and animal life; in some cases human health is adversely affected as well.

A common need of both the regulator and the regulated is to focus testing resources on those chemicals with the greatest potential for causing harm to man and the environment. Historically, the process of prioritizing chemicals for testing has included consideration of toxicological potency, estimated exposure, persistence, and bioaccumulation potential. The cost of generating information needed for even the simplest evaluations of the large numbers of chemicals (an estimated 45,000 existing chemicals and 1,000 new chemicals each year) is a national problem and has become the driving force for exploring relationships between chemical structure and those properties or effects considered essential for environmental hazard assessment.

The theoretical and practical considerations in the application of structure activity relationships (SAR) to chemical hazard assessment have been the subject of much research and debate. Most agree that SAR is a potentially valuable component of the early stages of chemical evaluation; research on SAR methods has received strong support from the Office of Research and Development (ORD) and the Office of Toxic Substances (OTS). In order to focus this research on producing the most cost-effective methods of chemical evaluation within the next few years, EPA has determined that now is an appropriate time to evaluate the current state of the art and to develop research programs for the next two years.

A workshop on "Structure Activity Concepts in Environmental Sciences" was held February 1-4, 1981, in San Antonio, Texas. Experts from industry, academia and government were invited to: (1) examine the state of knowledge of SAR for evaluating environmental fate and effects processes, (2) state the applicability and limitations of SAR predictive methods in regard to EPA's mission, (3) identify gaps in current knowledge and techniques, and (4) develop specific objectives and research priorities for the development of SAR over the next two years. The experts were organized into five working groups:

I. Available Methods for Structure Activity: Requirements and Limitations

II. Evaluation of Molecular Descriptors

III. Structure Activity Relationships for Environmental Fate Assessment

IV. Predicting Environmental Effects

V. Structure-Activity Application in Regulation

Each group produced a report summarizing the major points of discussion during the workshop, any conclusions reached, and specific recommendations regarding the EPA research program. Each working group report is presented in its entirety in the third Section of this document. Summaries of the general discussion contained in these reports follow immediately and the conclusions and recommendations of each group are presented in the next section.

The objectives of this study are (1) to examine the state of knowledge of SAR for evaluating environmental fate and effects processes (2) to indicate their scope and limitations and (3) to recommend needed research to extend the usefulness of SAR for regulatory purposes.

Group I - Available Methods for Structure Activity: Requirements and Limitations

This section describes mathematical and statistical methods that are used to approach QSAR problems. The methods are independent of the application area; thus, a given method can be used for a variety of studies, or a given application area can be investigated by a selection of methods.

The mathematical methods used to develop relationships between descriptors and responses constitute only one part of the overall process of SAR research. Results from mathematical analyses can be used to guide hypothesis formation, further descriptor development, further experimentation, further biological testing, and improvements in methodologies. Such feedback may lead to improved understanding of the problem under investigation.

Group I assumes the existence of a data matrix which may include molecular structure descriptors, physicochemical parameters, and/or biological responses. Solid SAR models can be developed using these mathematical and statistical methods. However, the validity of the models is contingent upon the quality of the elements of the data matrix. Therefore the application of SAR to problems of importance to EPA is dependent on the availability of suitable data and support from EPA. As the SAR community gains experience in analysis and expands the application range, some new mathematical methodology may be required.

The statistical method blends data analysis and probability theory. Important goals of the use of statistical analysis are to assess the variation in data and to arrive at probability-based summaries such as confidence levels and levels of significance of apparent relationships.

A relevant example is the task of predicting endpoints or parameters (Enslein and Craig 1978; Cramer 1980). This classical statistical problem uses multiple regression analysis, linear discriminant analysis, and similar (least-squares) statistical methods to explain the behavior of a set of dependent variables, such as LC50 or vapor pressure observed on a set of chemicals, in terms of a set of independent predictors. In a QSAR setting, the result is a mathematical equation that provides an estimate of the variable for untested chemicals.

Pattern recognition is a general name for a set of methods used to investigate data represented as points in n-dimensional spaces. There is a great deal of literature describing the field, and many of the techniques have been applied to structure-activity relationships problems.

An assumption in pattern recognition is that two points that are close together in n-dimensional space represent two compounds that are similar with respect to the molecular structure descriptors used. In the same way that two compounds that are close together are similar, a group of points forming a cluster defines a class or subclass of compounds.

There may be missing entries in the data matrix of any of the methods discussed. There is abundant literature on statistical methods for analyzing an incomplete data matrix, such as missing data regression analysis, discriminant analysis, and factor analysis (Beale and Little 1975; Rubin 1976; Hamilton 1975). Most statistical computer packages contain algorithms for treating missing data, but there is no standard algorithm, and some have been severely criticized (Heiberger 1977).

The primary limitations of analyzing a data matrix are the requirement for fairly strong assumptions about the statistical behavior of the data and the reasons for the missing values. In any particular application, it is therefore important to know what forms the basis of the method's assumptions and to test the data for their validity.

The methods used are differentiated between parametric and nonparametric. Parametric methods are those in which the data are assumed to follow a particular probability model, usually a model uniquely represented by mean vectors and covariance matrices. These include the typical statistical methods of regression, discriminants, and principal components analysis, as well as some pattern recognition methods. Caution in using such parametric measures should be exercised since they may not be justified for a particular data set. With nonparametric methods, the data themselves are used in the development of such methods as mappings, clusterings, and discriminants. The two categories are, however, not entirely disjoint.

SAR methods are information tools that should be used for internal priority setting within EPA. These methods must be embedded within the overall approach toward decisionmaking procedures. SAR predictions should not be the sole factor on which regulatory decisions are based.

Group II - Evaluation of Molecular Descriptors

The molecular descriptors used in QSAR have, historically, included many types not calculated directly from structure. The first such descriptor was, in fact, a parameter based on a physical property: the partition ratio of the solute between an immiscible lipid phase and water. Next on the scene came the "extra-thermodynamic" electronic parameters pioneered by Hammett (Hammett 1935, 1970) followed by steric parameters introduced by Taft (1956). Although both types have as their primary data source rate and equilibrium constants, some encouraging progress has been made in substituting for Taft's E_s a set of steric constants derived from van der Waal's radii and thus available directly from structure (Verlopp et al. 1976; Charton 1975).

Topological descriptors, such as molecular connectivity (Randic 1975; Kier and Hall 1976), are structurally derived but would be of limited utility without incorporating in the topology of each molecule some higher level of information about the nature of any heteroatoms it may contain. Valence factors derived from the number of valence and nonvalence electrons or else from the familiar field, resonance and molar refractivity parameters, can be added to accomplish this end (Yalkowsky 1980). Of course, the ultimate in understanding would be achieved when atomic structure is known in such detail that essentially all molecular properties, both chemical and physical, could be predicted prior to synthesis and testing. Quantum chemistry, based on the Schroedinger equation, has provided at least the beginnings of an approach to this theoretical ideal. Various semiempirical approaches have yielded useful information for example, on conformation, ionization potentials, and energy levels of activated intermediates (Kier, 1971).

From the foregoing it should be apparent that the subject of molecular descriptors can be categorized in several ways based on several perspectives. A historical development of descriptors also traces a trend from the more pragmatic to the more theoretical; from a dependence upon measurements of physical properties to a greater (or total) reliance upon involved computer calculation. The descriptors include a) those derived from physical properties, e.g., partition coefficient, molar refractivity, hydrophilicity, solubility; b) those based on reaction rate or equilibria and used in linear free energy relationships, e.g., Hammett sigma and Taft C_s; and, c) those calculated from structure, e.g., connectivity, upsilon, molecular volume and cavity area, structural 'keys', and quantum chemical indices.

Group III - Structure Activity Relationships (SAR) for Environmental Fate Assessment

Group III considered SAR for the major processes and properties that control fate. Although emphasis is on the use of SAR for estimating thermodynamic and kinetic constants, in many cases products of transformations can be predicted quite accurately (as for example in hydrolysis).

At present, it is Group III's belief that the following processes can be important in soil, water, and air:

SOIL/SEDIMENT	WATER	AIR
biotransformation	metal complexation	oxidation
hydrolysis	hydrolysis	photolysis
surface photolysis	oxidation	deposition
volatilization	photolysis	
sorption	sorption to sediments	
reduction	biotransformation	
bio-uptake	bio-uptake	

Movement among compartments is controlled both by equilibrium and kinetic processes, but estimates based upon thermodynamic properties can give a useful first approximation to distribution that will assist regulators and manufacturers in focusing attention on those compartments where a significant fraction of the chemical would accumulate. This distribution estimate provides only a boundary condition in many instances. More detailed estimates which include kinetic parameters will indicate that nonequilibrium conditions actually apply as for example in movement from soil or water to air.

The underlying assumption in use of kinetic and equilibrium is that the net movement and loss of a chemical is the simple sum of all such processes. If input is continuous, the net concentration of chemicals at any time will be in a steady state resulting from the balance of input and loss rates. Therefore, decisions about what constitutes a significantly fast or slow rate of loss must be tempered by consideration of the rate of input; for example, a chemical which hydrolyzes rapidly (half-life of a few days) may still be present in an aquatic system at significant concentrations if the input rate is high.

SAR can be a valuable means of eliminating certain processes from a detailed fate assessment if the value of the estimated rate or equilibrium constant is so small that the process cannot compete with other faster processes. Under these circumstances, the precise value of the rate constant is not important. Processes which have half-lives of many years or many hundreds of years generally will not control fate, and more reliable estimates of the rate constants are unnecessary. Thus, use of both qualitative and quantitative SAR can greatly simplify fate testing procedures.

Group III did not address the processes of uptake by plants of chemicals from soil or water, dry deposition of chemicals from the atmosphere, or reduction in soils/sediments. They concluded that all may prove to be important mechanisms for concentration or movement, but none appears to be well enough understood at this time to provide any basis for SAR.

Structure activity relationships (SAR) can greatly simplify the task of estimating rate and equilibrium constants needed for modeling transport and transformation processes in water, air, and soil. SAR can also assist in selecting best values from experimental measurements. Because the predictive power of SAR varies widely (factors of two to orders of magnitude), clear indicators of their accuracy and limitations are essential to avoid misapplication and unreasonable extrapolations and expectations.

The best developed SAR are for transport and chemical transformation constants in water and air. Thus solubility, sorption, and diffusivity can be estimated quite precisely from structural parameters. Hydrolysis in water and oxidation in water and air are well-understood processes, but SAR for these processes are limited by inadequate data bases. Metal complexation by natural and synthetic ligands are controlled both by thermodynamic and kinetic parameters; some SAR are available for metal ions and synthetic ligands; however, sorption and ion exchange with clay minerals may confound these estimates in unknown ways.

Photolysis in water and air often is complex, and the rate constants can be estimated only as upper limits. For biotransformations only broad generalizations and upper limits for rates can be given at this time.

Except for sorption, none of the processes believed to be important in soils (such as hydrolysis, biotransformation and uptake) are well understood and only very limited SAR are available at this time.

Group IV - Predicting Environmental Effects

Quantitative structure-activity relationships (QSAR) were viewed by Group IV as inadequte in present form for regulatory decisions except where scientifically based interpolations are possible. The working group agreed that development of QSAR:

- Provides a focused research target on a national problem and can be addressed effectively by many aquatic toxicological laboratories (government, industry, and academia) generating a minimum data base on a small but representative number of chemicals (rational groups)
- Is a cost-effective approach to help identify the spectrum of most-to-least hazardous chemicals but only after the necessary QSARs have been built and validated.

Until meaningful QSAR relationships are developed and validated, the historical process of prioritizing chemicals used by various academic institutions, industrial scientists and regulatory agencies should remain in practice. Specifically, these groups assign chemicals to a status of "low priority" based on various combinations of the following exclusion principles:

1. If the potential for exposure is low based on the environmental release rate
2. If the chemical is unstable in water based on chemical predictions,
3. If the chemical has been looked for and not found in the environment
4. If the soil-water-air equilibrium causes minimal concern, e.g., fugacity calculations based on structure
5. If the water solubility or molecular weight preclude significant accumulation, or
6. If in the wisdom of a panel of interdisciplinary experts the chemical is judged to be of low concern to man or the environment.

Since the objective of environmental QSAR development is to prioritize chemicals based on hazards, defined as the margin of safety between a no-effect concentration and an exposure concentration, this working group recommended no research and development or environmental decisions based on qualitative structure activity relationships, i.e., those based on pass-fail test results, or those with a questionable dose-response, or those with a nonquantitative dose-response, or those where the endpoints are not well defined, characterized, or are of unknown ecological significance.

The working group recommended the following Five Phase Research Program be initiated as soon as possible:

Phase I--Select about 1200 Chemicals, Divided into 60 Tests Sets, Representing All Relevant Chemicals.

The 1200 chemicals should be selected from a large list of chemicals including industrial chemicals, naturally occurring chemicals, drugs and pesticides which remain after a first-pass exclusion by assignment to "low priority" status. Also, the 1200 chemicals would include those "research" chemicals needed to complete a specific test set. Each test set should contain at least 20 chemicals which span a wide range of values described by molecular descriptors. The 1200 chemicals were chemicals/test sets multiplied by 60 test sets.

The selection of the 20 chemicals in each test set should be based on (a) chemical structure similarity, (b) at least 10 chemicals with "well-behaved functional groups" (characterized by weak hydrogen bonding) and 10 chemicals with functional groups of unknown chemical behavior and (c) chemicals with a similar mode of action according to the collective judgments and assumptions of medicinal chemists. For example, methanol and ethanol are similar chemically but might belong to different test groups because their mode of action is different.

Phase II--Generation of QSAR Database: Four Ecotoxicologically Meaningful Endpoints.

A. Lethality to aquatic vertebrates measured with fathead minnows in the 96 hour LC test.

B. Immobilization effects to aquatic invertebrates measured with Daphnia magna in the 48-hour EC50 test.

C. Chronicity factor (a modification of the conventional application factor), determined for several chemicals in each test set.

D. Lethality to mammals measured with rats in the seven-day LD50 test (diet or water).

These endpoints were selected for the following reasons:

1. Ecological importance--predictive relationships between these endpoints and other ecotoxicological data can be established. These other data include predicted lethal concentrations to a second aquatic species based on fathead minnow data, a predicted chronic no-effect concentration based on a chronicity factor and acute data, predicted field observation or aquatic epidemiological data based on laboratory data, and predicted no-effect concentrations to terrestrial animals based on the rat data.

2. Mammalian toxicology--In addition to the need to identify chemicals that cause harm to terrestrial animals, mammalian

toxicological QSAR were also included in the recommendation to emphasize the need and the mutual benefits of cross-fertilization between mammalian and environmental toxicology.

3. Standard test methods--The test procedures needed to generate this QSAR data base are reasonably well developed, but it was the intent of this working group for specific protocols to be drafted for each of these special QSAR tests to assure the highest quality data possible and to define the interlaboratory accuracy and precision. These protocols would specify several measurements which are not routinely made, including gross pathological observations for the most sensitive toxicological response and slope of the dose response and time to effect data. Also, protocols for water solubilities and octanol: water partition coefficients (at infinite dilution) were recommended, since correlations with these chemical properties are anticipated. These maximum effort protocols should not be considered necessary for testing specific chemicals for conventional hazard assessments in the future, but are necessary for building these QSARs for ecological extrapolations.

Phase III--Consolidation of the 60 Test Sets of Chemicals into a Minimum Number of Rational Groups.

The process of lumping and splitting chemicals into ecologically meaningful rational groups was seen by this working group as the largest challenge before using QSAR for prioritizing chemicals. This challenge should be greatly aided by computerized data analysis of the various combinations and permeations of molecular descriptors and toxicological endpoints to identify the most reliable regressions.

Phase IV--Refinements and Improvements

New molecular descriptors, new endpoints, confidence limits of predictions, accuracy criteria for the use of QSAR in environmental decisions, and validation of at least one benchmark chemical for each rational group of chemicals were areas needing refinement and improvement after the rational groups were identified. Many possible new endpoints could be evaluated including predictors of the rate of metabolic toxification and detoxification in aquatic organisms, time-independent ECi values (concentrations causing i per cent of some effect) and bioconcentration potential for chemicals not fat soluble.

The uncertainty associated with each of the various predicted values will get larger with each extrapolation. Therefore, it is important to assign confidence limits to each predicted value and to prepare an accuracy criteria for accepting or rejecting individual values. Comparisons of the values for untested new chemicals with those values of benchmark chemicals of similar structure improves the confidence of the scientist for a specific predicted value. Furthermore, if a new chemical fails to fit into one of the rational groups for using QSAR, the scientist can draw on other sources for wisdom to help place the new chemical in its proper priority.

Phase V--Role of QSAR in Hazard Assessment: Extrapolations of Endpoint Values to Ecological No-Effect Concentrations

In Figure 1 Group IV summarizes its viewpoint of the role of QSAR in the overall process of prioritizing chemicals. The process starts with a chemical structure and ends with an estimate of hazard and possibly an environmental decision. The key to this viewpoint was the need to estimate a toxicological endpoint value from regression equations based on calculatable descriptions of the chemical structure, i.e., molecular descriptors.

Specifically, this working group addressed the relationships between structure or physical properties and the toxicity to aquatic organisms. Because of its importance in toxic effects, bioaccumulation potential in

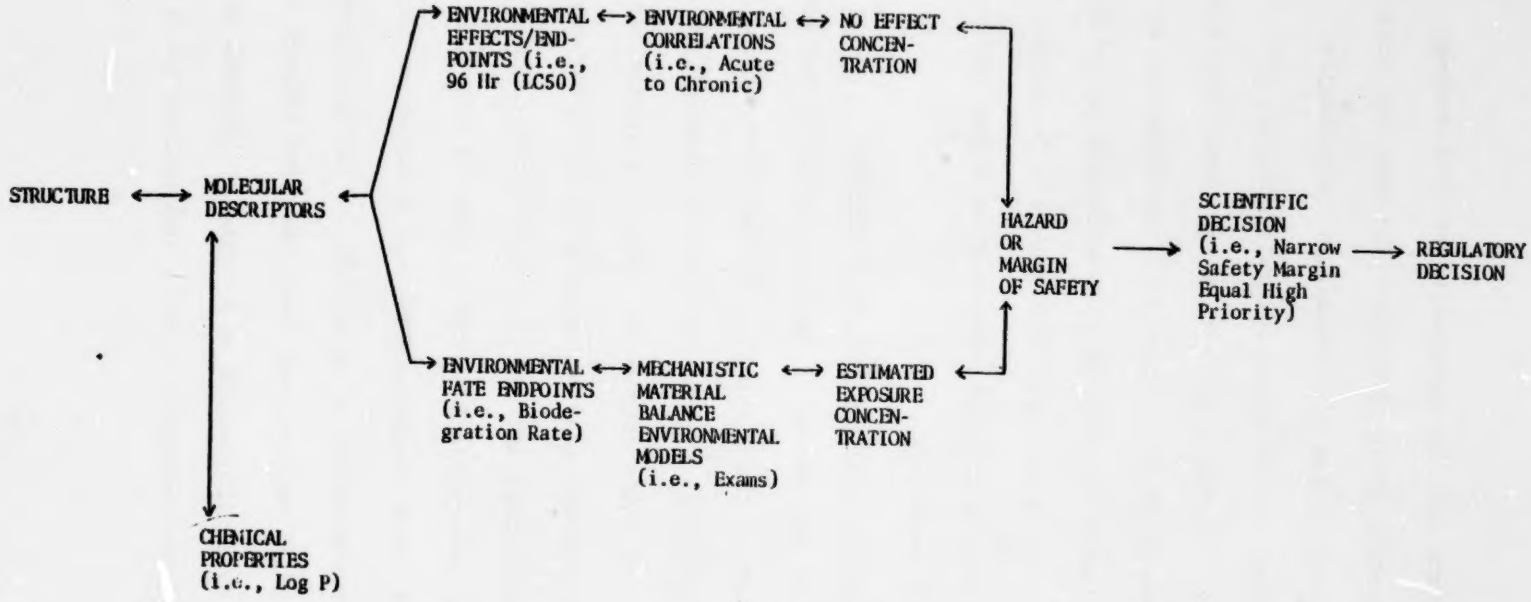

STRUCTURE
MOLECULAR DESCRIPTORS
CHEMICAL PROPERTIES (i.e., Log P)
ENVIRONMENTAL EFFECTS/END-POINTS (i.e., 96 Hr (LC50)
ENVIRONMENTAL CORRELATIONS (i.e., Acute to Chronic)
NO EFFECT CONCEN-TRATION
ENVIRONMENTAL FATE ENDPOINTS (i.e., Biode-gration Rate)
MECHANISTIC MATERIAL BALANCE ENVIRONMENTAL MODELS (i.e., Exams)
ESTIMATED EXPOSURE CONCEN-TRATION
HAZARD OR MARGIN OF SAFETY
SCIENTIFIC DECISION (i.e., Narrow Safety Margin Equal High Priority)
REGULATORY DECISION

FIGURE 1.

STRUCTURE TO HAZARD AND DECISIONS

aquatic organisms was also considered by this working group. The working group recognized the need for parallel reviews of relationships with mammalian toxicological effects including carcinogenicity, mutagenicity, and teratogenicity. It was also felt by the group that too little is known about the uptake and transformation of chemicals by plants to consider the applicability of QSARs at this time. It is intended that the relationships discussed in this section be used only as a part of an overall hazard assessment.

Group V - Structure-Activity Application in Regulation

The rapid growth in the number of chemicals used by our society and the late development of the scientific knowledge regarding possible adverse environmental impacts has left us with a lack of information about many of these chemicals. With the recognition of the lack of this knowledge has come the determination to develop a rational basis for manufacture and use to avoid unacceptable adverse consequences. This is demonstrated in the passage of the Toxic Substances Control Act (94th Congress 1976) and a number of other acts providing for the regulation of chemicals in the environment. These laws and promulgated regulations under the laws attempt to employ the best available information as a basis of judgment of safe manufacture and use of the chemicals. Thus, information as to the substance's chemistry, biological effects, method of transport, use, and disposal are always brought into regulatory consideration. However, the data base, particularly that of the chemistry in the environment and biological effects, often contains serious gaps. In part this is due to the methods that must be used in acquiring the information; for example, to evaluate the biological effects of a single chemical on a number of organisms at several different exposure levels would entail an effort of perhaps as much as five work years and several hundred thousand dollars. Similarly, the determination of the environmental fate and behavior would involve nearly as much time and effort. Thus it is, that for many of the 45,000 chemicals in commerce, little of this type of information has been developed.

It is important to note that not all chemicals pose the same magnitude of problems. In fact many, for example polymers, are relatively inert, causing little or very small biological effects. Likewise, their environmental impact can be readily minimized. Therefore, it would be highly desirable to have a scheme of categorization and prioritization that would enable us to distinguish between the chemicals posing few, if any problems, from those that are potentially troublesome. With such a scheme, the attention and resources could then be focused on the problem chemicals and the development of a rational procedure of manufacture, use, and disposal. It is on this basis that the concept of the hierarchy of testing came into being. Even so, if in the hierarchy of testing, long-term biological assays or monitoring are required, it still becomes difficult and time consuming to identify the more troublesome of the chemicals for further attention. For that reason, a more rapid method of identifying those chemicals is required.

In the nineteenth century, scientists investigating the biological activity of chemicals developed a concept that the active chemicals reacted with receptor sites in the living organism, and that in order to do so must have a spatial configuration and physical properties that accommodate to the receptor site. This was a "lock and key," Ehrlich. Subsequently, it was observed that with narcotic agents certain properties, notably water solubility, vapor pressure, and partition coefficient, were associated with that activity. Based on these early observations, the chemists and biologists postulated that there was a necessary and sufficient relationship between the chemical and physical properties of the compound and the biological effects produced (Sexton 1963; Albert 1973). A great deal of research effort has gone into attempting to elucidate an understanding of this relationship and generalize it to fit all instances where compounds exhibit biological activities. It was hoped that from this, the biological activity of a new structure could be predicted, thus facilitating both the synthesis of new compounds of utility and the development of appropriate methods of managing existing compounds.

Efforts have increasingly turned to devising quantitative structure-activity relationships by which one could predict the particular biological response that would be produced by a chemical. Similarly, this information would also be used in assessing probable environmental fate and behavior. The so-called quantitative structure-activity relationship (QSAR) in fact encompasses a broader concept than structure, inasmuch as it is concerned with both chemical and physical properties (Albert 1975).

The QSAR approach involving both chemical and physical properties has become increasingly sophisticated thanks to the efforts of a few dedicated investigators. From its early beginnings where attempts were made to relate simple elemental structure, molecular weight, vapor pressure, water solubility, and partition coefficient to biological activity, it has been elaborated to utilize the much more refined and sophisticated chemical concepts of kinetics and quantum mechanics. While the most successful applications have utilized the linear free energy approach such as that based on partition coefficient coupled with the Hammett constant, other approaches utilizing molecular orbital calculations, cluster analyses and pattern recognition have all served to increase the power of QSAR (Topliss 1978; Helmer 1968).

Also in the case of compounds possessing significant potential biological activity, appropriate biological and environmental investigations would be required to provide the refined answers needed. At the moment, Group V also believes that QSAR is not sufficiently developed to be more than an effective preliminary screening tool where experimental data are lacking. The greatest use of SAR will be ensured if the available information is compiled into a computerized data base. At the moment, QSAR's greatest value is in making these evaluations in the absence of empirical data.

In considering the priority problem of developing rational guidelines and regulations for the safe manufacturing and use of chemicals, it would appear that few techniques of evaluation offer more potential value than QSAR.

While newer techniques and approaches to evaluation may develop in the future, at the moment it would appear that this method offers sufficient promise that those applicable QSAR techniques should be utilized, and as the field advances further, the new development adapted as quickly as possible to the regulatory procedures. Indeed, it is felt in many quarters that at the moment, QSAR offers perhaps the major, if not the only means of quickly evaluating the environmental impact and possible biological effects of the host of chemicals employed by a technological society.

CONCLUSIONS AND RECOMMENDATIONS

Group I - Available Methods for Structure Activity -- Requirements and Limitations

- Mathematical methods are available to develop appropriate SAR models, discriminants, etc.
- Model validity is dependent on a good data matrix.
- As SAR community gains experience in analysis and begins to expand the range of application, new mathematical methodology may be required.
- SAR methods are informational tools.
    - Use should be for internal priority setting within EPA
    - SAR methods must be only a part of the overall approach toward decision-making procedures for regulatory decisions.

Recommendations

- Group I recommends that EPA take an active role in the organization and accessibility of structure activity data
- Participate in the support of centers of excellence in SAR for the development of expertise, including programs for training investigators.
- Support the generation of high quality SAR data bases
- Develop a clear set of procedures for the submission of grants in the SAR area, including the personnel directly involved.
- Provide long-term stable research support for SAR research in addition to short-term, contract support for specific goals.

Group II - Evaluation of Molecular Descriptors

Molecular Descriptor 1: Partition Coefficient

- Log $P_{o/w}$ has been shown by several investigators to be related to bioaccumulation (Neeley 1974, Veith   ). Its relationship to binding to the organic component of soil, K_{oc}, (Kenaga 1980) makes it valuable in transport models. Because it models nonspecific hydrophobic interactions, it can be a useful predictor of certain types of toxicity, such as narcosis, hemolysis and necrosis.

- It should not be expected that a linear relationship of the sort which is found with lower homologs will be maintained indefinitely.

Recommendations

- Listed in order of priority, the recommendations to the U.S. EPA for initial support are:
  - Comprehensive literature search and data base maintenance and design; measurement by HPLC of very lipid solutes;
  - Further testing of the structural limitations to the relationship of log $P_{o/w}$ to log K_{oc} and log k_1/k_2 (bioaccumulation).
- Continued support is recommended for:
  - Measurement by standard shake-flask procedure of solutes whose structures will add new fragment values for CLOGP
  - Improvements in the CLOGP algorithm itself.

Molecular Descriptor 2: Solubility

- A knowledge of solubility is essential for understanding and predicting the fate of compounds in the aqueous environment because it can be a limiting parameter in the bioaccumulation of organic materials.
- Solubility can be used to estimate the time course of pollution resulting from a single spill and the amount of a compound that can be carried away by a river.

Recommendations

- It is recommended that EPA fund a coordinated effort to obtain reliable aqueous solubility data through literature search and measurement.

- This group also recommends that EPA support research in correlating solubility with other physical and biological properties.

Molecular Descriptor 3: Molar Refractivity

- This descriptor primarily relates to molar volume but has a component of polarizability that has been shown to be useful in modelling dipole-dipole interactions or binding due to London dispersion forces.

- It appears to be uniquely useful in correlating solute binding to (or inhibition of) isolated enzymes—an important factor in toxicity evaluations.

- Molar refractivity has not yet been established as an important parameter for environmental (QSAR).

    - Sole effort has been to calibrate halogen valences in the molecular connectivity approach (Kier 1976).

Recommendations

- Although a high priority is not recommended, this group suggests a very modest level of financial support to assure:

    - Expansion to several thousand of the data base of structures for which Lorentz-Lorentz calculation could be made.

    - Compatibility of the computer calculation program with EPA's facilities and its availability in the shortest possible time.

    - Investigations are made into the separate contributions of the components of molar volume and polarizability to determine if the combination in present use is optimal.

Molecular Descriptor 4: Hydrophilicity

- In order to determine the hydrophilicity of a molecule, its vapor pressure should be large enough to measure. If the vapor pressure is measurable, there will essentially always be enough water solubility, but occasionally compounds will react too rapidly with water and the hyrophilicity will not be measureable.

- The correlation in terms of molecular structure is good for monofunctional compounds but is sometimes poor for polyfunctional compounds.

- Hydrophilicities are directly relevant to the following types of EPA questions:

  1) What is the tendency for a compound in aqueous solution to go into the vapor phase; i.e., from a river, lake, or ocean to the air; or from the wet surface of lung into the air, or the reverse?

  2) What is the tendency for a compound to go from the vapor phase into aqueous solution; i.e., how well will rain wash a pollutant from the air, or how much inhaled pollutant will go into the moist surface of a lung?

  3) The correlations in terms of molecular structure permit one to estimate the vapor pressure of a compound from its water solubility or to estimate its water solubility if its vapor pressure is known.

Recommendations

- EPA should support a comprehensive literature search and measurements where necessary to determine more hydrophilicity values in order to increase the representation of certain structural features that are common to industrial chemicals and to biologically active compounds. This should permit a substantially improved understanding of the proximity effects--the deviations from "ideal" behavior shown by polyfunctional compounds.

Molecular Descriptor 5: B,C Parameter

- A consensus exists that today's QSAR methodologies are more likely to help in the definition of the concentration of pollutant at some locale, rather than the toxicological implications of that concentration. BCDEF interrelationship might help predict the behavior of unknown compounds in either of two ways:

  1) If kinetic or physical models governing this fate of molecules in the environment can be established, then the BCDEF interrelationship might define the parameters which control the behavior of an unknown in this model; e.g., binding to clay, vaporization from water, partitioning into fatty substances.

2) Failing the existence of such explicit kinetic models, it may be possible to define and measure the concentration at some location, 'B', as being at the threshold of environmental risk and as a function of the chemical leaked at some other point, 'A', and carry out this experiment for twenty other compounds chosen from the list given in J. Am. Chem. Soc., 102, 183 (Cramer 1980). It is reasonable to hope that the BCDEF parameters will model this process, independent of its mechanistic detail.

- If neither condition (1) or (2) define endpoints of interest to EPA, then it is difficult to see how any QSAR approach can help, since all QSAR methodologies rely on fitting parameters to end points.

Recommendations

- The pioneering work with the BC-Parameter was undertaken at a drug company to clearly separate nonspecific biological effects from those which might be optimized to yield a marketable product. It has succeeded to the extent planned, and it is unlikely that this study will be continued privately. Since the separation of nonspecific toxicity from unique toxicity is of high priority to EPA, and the BC-Parameter promises to be very helpful in this application, a rather high priority should be assigned to its support.

- In the minimal set of properties necessary for BC-Parameter calculation, measurements of some are already receiving support. Support for others should be considered as well as support for their inclusion in an accessible data base.

Molecular Descriptor 6: X-T Paramters

- The acceptance of LFER QSAR methods in drug and pesticide design is a result in part of the method's simplicity and economy of time and money. Much of this has been accomplished using X-T parameters of the types discussed in this section, but, on the other hand, many other efforts were thwarted for lack of values. The task of hazard assessment facing OTS encompasses a much broader class of chemicals and thus requires a much more extensive data base than that barely adequate for the drug and agrochemical industries.

Recommendations

- It is recommended that at least medium priority be given to support of X-T parameter measurement (e.g., there is a real need for more σ_R^+ and σ_R^- values), and for a critical compilation of those reported in the literature with special attention to estimates of error in these values.

Molecular Descriptor 7: Degree of Ionization

- It would seem that the degree of ionization, derivable from the pKa, which is itself an X-T parameter, ought to be included in the previous section, and, in fact, ought to be adequately covered when the appropriate sigma values are used. This has not proved to be the case in actual practise, however.

- The method of pK_a calculation presented by Perrin (Yalkowsky 1980)[a] provides a good foundation, but there is room for improvement, for the error is ± 0.4 in many cases. This effort would be aided materially if there were a comprehensive collection of pK_a's in computer-readable form.

Recommendations

- Establishment of a data base, searchable down to connection tables, of existing pK_a data.

- Research on methods to calculate pK_a' from structure at least comparable to those now available for conformation and log $P_{o/w}$.

- Research on how to treat ionization within the context of QSAR. Some drug research efforts in this area may have implications for EPA problems; e.g., the efforts to change urine pH to help the body eliminate an overdose of a drug (acid pH for bases such as amphetamines and alkaline pH for acids) may suggest some "ion-trapping" procedures of use to deal with environmental hazards.

Molecular Descriptor 8: Molecular Connectivity

- Molecular connectivity indexes have been used in three areas which are directly related to EPA needs:

1) Physical properties such as solubility, volatility and $P_{o/w}$ are necessary in environmental fate models, and connectivity has been useful in estimating such properties directly from structure.

2) In the literature there are some toxicity correlations which have used connectivity. ERL-D has supplied LC50 data on the fathead minnow have been used to derive correlation equations. More studies of this type can provide a base line for general narcosis and a means of identifying suspected highly toxic materials.

3) The structural information contained in connectivity indexes may be used with pattern recognition techniques to help classify highly toxic substances.

Recommendations

- Connectivity indexes can be made more readily accessible to the QSAR community on a broad basis. The computer programs which are already available can be distributed or access encouraged to a central data base, or a combination of the two.

- In cooperation with EPA laboratories, the limits of the indexes should be explored by application to a wider variety of data sets. Specific problems should be defined, data selected from the literature or obtained from experiment where necessary, and model studies performed. Predictive ability should then be tested by considering data not included in the original data set.

- Classification schemes should be investigated which are intended to give broad indications of toxicity or level of effect. Predictive equations can be developed to estimate magnitude of effect.

Molecular Descriptor 9: Substructural Key Systems

- Various substructural key systems have been used in SAR, including CIDS and CROSSBOW. The original 148 CROSSBOW keys which are based on Wiswesser Line Notation (WLN), have been augmented by further descriptors, including those substructures considered as suspicious of carcinogenicity by FOA Bureau of Foods. The present list thus consists of about 350 keys.

- Another set of keys is under development which will, it is hoped, be more meaningful chemically than the WLN-based keys. These keys are based on a system developed sometime ago at Smith, Kline and French. With appropriate support it would be possible to input these keys into a graphics system that would also be compatible with the Chemical Abstracts connection tables.

- Encouraging results (Enslein) have been obtained using these keys in analyzing rodent toxicity and carcinogenicity data. The chances are good, therefore, that they would prove effective in studying a wide variety of environmental hazards.

Recommendations

- EPA support is recommended for the purchase by qualified reseachers of graphic systems which would permit greater development of SAR methods applied to toxicology.

Molecular Descriptor 10: Molecular Volumes and Areas

- Many of the descriptors needed to develop QSAR for enviromental applications require the measurement or estimation of certain physical properties. There is general agreement that, for the near future, most of these will have to be calculated from molecular structure. The initial success of surface area and volume based approaches to the prediction of solubilities, partition coefficients and dissolution energies suggest that it could become, at the very least, a valuable confirmation for values obtained by other procedures.

- To avoid needless duplication of effort, it seems appropriate to use the newly developed software to compile a data base of surface areas and volumes for molecules of special interest to EPA.

- These more accurately computed areas and volumes should be used to construct a table of substituent group areas and volumes analagous to that published by Bondi (1964).

Recommendations

- Although support for further development of software for the computer calculation of solute areas and volumes cannot be placed on a high priority basis, the EPA should encourage further studies relating these descriptors to others which have been used earlier in QSAR.

- It should consider making the current program available to its regional laboratories which carry out modeling and toxicity studies so that these descriptors can be conveniently included in the early development of QSARs. Only in this way can it be determined whether or not solute area and/or volume contain information missing in the descriptors in more common use.

Molecular Descriptor: 11: Descriptor Reduction

- The availability of structural factors for a large number of compounds can have a number of main impacts which are in EPA's interest:

  - The very existence of this set will promote QSAR work on the chemicals included in the set. Assuming that important industrial chemicals are included, much of the QSAR output based on the set could be directly used by EPA.

  - Since the original descriptors are calculated solely from the molecular structure, predictions for many new chemicals could be rapidly made.

  - QSARs designed for screening purposes, which are based on the structural factor set, are of particular EPA interest. These screening methods could be utilized without waiting for lengthy lab tests. It should also be noted that these methods would be useful to industry. Chemicals could be "prescreened" for likely adverse effects before extensive efforts have gone into synthesis, production, and marketing.

Recommendations

- It is recommended that EPA support an effort to compile or compute molecular descriptors which can be calculated from molecular structure, where these methods exist, and to encourage development of such descriptors The reduction of such descriptors should be done by PCA and the resultant factors should be disseminated to the QSAR research community.

Molecular Descriptor 12: Theoretical Chemical Descriptors

- One of the least tractible problems in hazard assessment is that of a chemical which only becomes toxic after a metabolic process is initiated. If the actual toxicant appears only after several steps in the process, the molecular descriptors information obtained from the starting material is often of little or no relevance. Even though very few mechanistic studies have been completed in detail for metabolically-activated toxicants, there is already evidence that the action of cytochrome P-450 can be involved in both acute and chronic toxicity.

Recommendations

- Obviously EPA cannot support in-depth mechanistic studies of all potential toxicants regardless of how desirable such knowledge would be. But when an ubiquitous enzyme, such as cytochrome P450 is implicated in a number of metabolic activations leading to toxicity, then support of efforts to utilize these findings in SAR compatible to, and supportive of, other ongoing QSAR efforts is certainly recommended.

Group III - Structure Activity Relationships (SAR) for Environmental Fate Assessment

Environmental Fate Process 1: Photochemical Reactions of Organic Chemical in the Environment

- To relate structure to activity is enormously more difficult in photochemical reactions than it is in thermal reactions.
    - The reactant in the photochemical process is the electronically excited state of the molecule which can be looked upon as a new, transient chemical compound, whereas the structure examined is that of the ground state.
- Attempts to extend quantitatively the relationship between structure and photochemical reactivity in a series of homologous compounds under controlled conditions (wavelength of light, temperature, solvent) have resulted in very limited correlations.

- Practical problems in the use of SAR also exist.
  - Chemicals in the environment are found as dilute solutions in water and air.
  - Data on both absorption spectra and photochemical processes rarely extend to water or the gas phase.
  - Extrapolation of the data from one solvent medium may be unwarranted.
- Data on the photochemistry of organic compounds in aqueous media and in air are quite limited, although most relevant and essential for the present purpose.
- Organic chemistry has grown extensively over the past 25 years, but there are still large gaps in our knowledge that make the prediction of the reactivity and the nature of the product for many compounds essentially a guess.

Environmental Fate Process 2: Metal Complexation and Speciation

- An organic chelating agent released into environmental waters may undergo extensive changes of toxicity (usually reduction of toxicity) as its speciation changes from that of a free ligand to the chelates of the naturally occurring metal ions. Toxicities of chelating agents and their metal chelates may differ by several orders of magnitude, and the chelates of metal ions that are naturally present in biological systems generally have the lower toxicities.
- Similar changes in the toxicities of metals with speciation, which vary widely with human dosage or exposure level, have also been observed. Administration of many metal ions at trace concentrations is not harmful since they would be complexed by naturally occurring ligands to form metal transport complexes and metal-activated enzymes. When these metals are supplied in excessive amounts, however, the natural metal receptors are overwhelmed, and the excess metals may then have toxic effects on other biomolecules that have coordinating donor groups.
- Metal complex speciation may change drastically with concentration in both environmental waters and in biological fluids, because of competition with low but often significant levels of naturally occurring metal ions and ligands. It is generally not possible to extrapolate toxicity data measured

at high concentrations to low environmental or biological levels. This behavior contrasts to that of organic molecules that do not change composition on dilution, and for which the extrapolation of toxicity data to low concentrations is sometimes considered valid.

Environmental Fate Process 3: Hydrolysis in Aquatic Systems

- Investigation of the rates and mechanisms of hydrolysis reactions has been a major research area in physical organic chemistry for several decades, and the detailed mechanisms that are operative in mixed aqueous-organic solvent systems are reasonably well understood for a variety of organic substrates. SARs in the form of linear free energy relationships (LFErs) have been developed for several categories of compounds including carboxylic acid esters, amides, acyl halides, benzyl halides, phosphate esters, carbamates, and others (Chapman and Shorter 1972; 1978; Lessler and Grunwald 1963; Wells 1968). In recent years attention has focused on determination of rate constant and development of LFERs for hydrolysis in purely aqueous solvent, to generate relationships with more predictive value for aquatic environments.

- Qualitative SARs are useful estimators of relative rates of hydrolysis under similar conditions.

- A minimum condition for confirmation of the constancy of mechanism is that formation of analogous hydrolysis product(s) from each reaction in a series used to develop an LFER should be demonstrated.

- LFERs with broad applicability seem unlikely to be developed for hydrolysis reactions. Even for such similar series of reactions as hydrolytic displacement of substituted phenoxide ion from phenyl carbamates vs phenyl n-methyl carbomates, Wolfe et al., (1978) found significantly different regression lines in plots of k_h vs pK_a of the leaving group.

Environmental Fate Process 4: Oxidation in Water

- SAR for $^1O_2$ oxidations are very limited. Foote and Denny (1971) measured the relative rates of oxidation of a series of substituted styrenes and found that for allylic oxidation correlation by the Hammett equation was satisfactory. However, epoxidation of the styrene double bond which accompanies allylic oxidation is not correlated well by nor $^+$, but is correlated by $[\sigma - 0.37\ (\sigma^+ - \sigma)]$ to give $\rho = -0.87$.

Environmental Fate Process 5: Biotransformation Processes

- The general pathways for microbial transformation and mineralization of many common organic compounds are well known. Substituents that interfere with these recognized pathways often tend to make the molecules more resistant to transformation. However, few quantitative relationships relating structure and activity have been developed.

Environmental Fate Process 6: Fate of Chemicals in the Atmosphere

- The physical and chemical processes that affect chemicals emitted into the atmosphere are reasonably well understood.

- Rates of these processes can be readily treated by various prediction methods.

- The physical processes:

  1) Dilution by diffusion

  2) Rainout and washout

  3) Dry deposition

  4) Sorption to aerosol particles
  are believed to be of minor importance in determining the fate of organic compounds.

- The fate of a large majority of compounds is determined by the following three chemical processes:
  1) Reaction with OH radical

  2) Reaction with ozone

  3) Reaction by photolysis

- SAR methods are recommended for the three dominant reaction processes.

  - First, for the reaction of OH radical, the procedure is based on the group additivity principle (Benson, 1976).

The method identifies the reactivity of various primary groups, such as primary, and tertiary carbon hydrogen bonds, and the effects that various substitutents have on these reactivities.

- The estimation of ozone reaction rate constants is limited to reactions with olefins and highly substituted aromatic compounds. One estimates the k values by matching the structure of the pollutant with similar structures of compounds with known reactivity. Care must be taken in determining reactivity of compounds with nonhydrogen substitution. For most compounds this method is expected to give rate constants with an accuracy of a factor of ±3. (Hendry 1979, 1980).

- The proposed method for analyzing photolysis rate is to calculate the rate of absorption of light on a clear day, usually the summer solstice. Then a maximum possible rate of reaction is estimated by assuming a quantum yield for chemical reaction of unity. In most cases this maximum limit is much slower than the rate with OH or O_3. Thus it is possible to eliminate the possibility of photolysis being a chemical process. In cases where the maximum limit rate is comparable or greater than the sum of OH and O_3 reaction, then it is not possible to say how important photolysis will be and an experimental determination of the quantum yield will be necessary.

Environmental Fate Process 7: Transport of Chemical Substances in the Environment

- For assessing the environmental fate of chemical substances several classes of chemical transport property data are required in addition to the usual information about the magnitude of the sources, the receiving environment and the prevailing environmental conditions that include reactivities, temperature, and wind and water flow.

- These classes are:
  - Equilibrium Partition Data Between Media
  - Transport Rate Terms Between Media

- SAR has been a useful tool for developing phase equilibrium information on aqueous solubility, Henry's Law Constant, octanol-water partition coefficient, sorption partition coefficient, and bioconcentration factor. There is, at present, inadequate use of SAR for estimating vapor pressure.

- Adequate correlations are available for diffusivities in air and water as a function of the molecular volume or weight (Reid et al. 1975). There is thus little potential for SARs to contribute further to transport rate prediction. The major problem is the characterization of the turbulence near interfaces.

Recommendations

Group III identified the following major research needs as essential for additional development of useful and reliable SAR for environmental processes; needs are ranked somewhat by overall assessment of importance and likely early progress.

- Biotransformation in Water or Soil - Develop a better understanding of the molecular basis for acclimation and transformation, including hydrolysis, oxidation, and reduction in order to more accurately measure rate constants for biotransformation at low concentrations of chemicals, and from this information develop SAR for the reactions.

- Soil Processes - We need to understand how soil humus and clay minerals interact with chemicals, and can either catalyze or inhibit hydrolysis, oxidation reduction, photolysis, and volatilization in order to provide the mechanistic basis needed for development of SAR.

- Hydrolysis in Water - Extend the kinetic data base for hydrolysis in pure water for more classes of structures over a useful pH range (3-11); explore structure-mechanism relations in water and application of existing SAR developed in mixed solvents.

- Photolysis in Water and Air - Extend the data base for quantum yields for photolysis of structurally similar compounds; explore how UV spectra and quantum yields measured in nonpolar solvents might be used in water or air.

- Oxidation in Air - Extend additivity SAR for oxidation by HO radical to more substituted alkanes, olefins and aromatics; explore development of SAR for oxidation by ozone.

- Oxidation in Water - Develop data base for oxidation of reactive structures by RO_2 radical and singlet oxygen in water; examine the applicability of existing SAR based on reactions in organic solvents.

- Volatility - Develop SAR for vapor pressures of high molecular weight chemicals based on critical volumes and boiling points.

GROUP IV - Predicting Environmental Effects

- There was general agreement that ecologically meaningful relationships between chemical structure and biological activity can and should be developed as tools for the prioritization of chemicals according to their degree of hazard in the environment.

- QSAR are inadequate in present form for regulatory decisions except where scientific interpolations are possible.

- Development of QSAR provides:

  - A focused research target on a national problem that can be addressed effectively by many aquatic toxicological laboratories generating a minimum data base on a small but representative number of chemicals (rational groups)

  - A cost-effective approach to help identify the spectrum of most-to-least hazardous chemicals after the necessary QSARs have been built and validated

- Since the objective of environmental QSAR development is to prioritize chemicals based on hazards, defined as the margin of safety between a no-effect concentration and an exposure concentration, this working group recommended no research and development or environmental decisions based on qualitative structure activity relationships, i.e., those based on pass-fail test results, or those where the endpoints are not well defined, characterized, or are of unknown ecological significance.

Recommendations

- The following Five-Phase Research Program should be initiated as soon as possible

  Phase I--Select about 1200 chemicals, divided into 60 test sets, representing all relevant chemicals

  Phase II--Generation of QSAR database: Four ecotoxicologically meaningful endpoints

1) Lethality to aquatic vertebrates measured with fathead minnows in the 96-hour LC50 test

2) Immobilization effects to aquatic invertebrates measured with Daphnia magna in the 48-hour EC50 test

3) Chronicity factor (a modification of the conventional application factor), determined for several chemicals in each test set

4) Lethality to mammals measured with rats in the seven-day LD50 test (diet or water)

Phase III--Consolidation of the 60 test sets of chemicals into a minimum number of rational groups

Phase IV--Refinements and improvements of:

1) New molecular descriptors, new endpoints, confidence limits of predictions, and accuracy criteria for QSAR use.

2) Validation of at least one benchmark chemical for each identified rational group of chemicals.

Phase V--Role of QSAR in hazard assessment: Extrapolations of endpoint values to ecological no-effect concentrations

1) Until some experience is gained in predicting effects and fate of chemicals by QSAR, regulatory decisions should not be based solely on QSAR predictions.

2) It is recommended that existing regulations be evaluated by QSAR predictions to gain experience in the process.

Group V--Structure-Activity Application in Regulation

- Few techniques of evaluation are of more potential value than QSAR for the priority problem of developing rational quidelines and regulations.
- Applicable QSAR techniques should be used for evaluation.
- New QSAR developments should be adapted as quickly as possible to regulatory procedures.
- QSAR's greatest value is evaluating the environmental impact and possible biological effects of chemicals in the absence of empirical data.

- Compounds with significant potential biological activity require appropriate biological and environmental investigations.

- QSAR is sufficiently developed only as an effective preliminary screening tool where experimental data are lacking.

Recommendations

- Available information should be compiled into a data base.

- EPA should encourage and support increased research and development of SAR, a general and reliable technique of evaluating chemicals for regulatory purposes.

REPORTS OF WORKING GROUPS

GROUP I

TABLE OF CONTENTS

Page

GROUP I - AVAILABLE METHODS FOR STRUCTURE ACTIVITY - REQUIREMENTS AND LIMITATIONS

1.0 INTRODUCTION 1

2.0 DATA 2

2.1 Statistical View 2
2.2 Pattern Recognition View 3
2.3 Analysis in the Presence of Missing Data 5

3.0 METHODS 7

3.1 Regression 7

  3.1.1 Data Requirements 8
  3.1.2 Availability 9
  3.1.3 Limitations 9
  3.1.4 Multivariate Considerations 11

3.2 Discriminant Analysis for Categorical Endpoints 12
3.3 Factor Analysis/Principal Components Analysis (FA/PCA) 12
3.4 Discriminants 15
3.5 Cluster Methods 17
3.6 Mapping and Display 18

4.0 APPLICATION TO U.S. E.P.A. PROBLEMS 21

5.0 RECOMMENDATIONS REGARDING RESEARCH OPTIONS 22

REFERENCES 23

45

GROUP I

AVAILABLE METHODS FOR STRUCTURE ACTIVITY - REQUIREMENTS AND LIMITATIONS

1.0 INTRODUCTION

This section describes mathematical and statistical methods that are used to approach QSAR problems. The methods are independent of the application area; thus, a given method can be used for a variety of studies, or a given application area can be investigated by a selection of methods.

The mathematical methods used to develop relationships between descriptors and responses constitute only one part of the overall process of SAR research. Outputs from mathematical analyses can be used to guide hypothesis formation, further descriptor development, further experimentation, further biological testing, and improvements in methodologies, etc. Such feedback may lead to improved understanding of the problem under investigation.

We assume the existence of a data matrix which may include molecular structure descriptors, physicochemical parameters, and/or biological responses. While we recognize that the quality of the elements of the data matrix are crucial to the overall success of any SAR study, this section does not address this aspect of the SAR problem.

2.0 DATA

Sections 2.1 and 2.2 describe the statistical and pattern recognition vocabulary used to describe the data analyzed and the mathematical methods used. The two views overlap and should not be considered mutually exclusive. A given problem can often be formulated using either approach, perhaps even using equivalent mathematics.

2.1 Statistical View

The statistical approach has traditionally been a blend of data analysis and probability theory. Important goals of a statistical analysis are to assess the variation in the data and to arrive at probability-based summaries such as confidence intervals and levels of significance of apparent relationships.

A relevant example is the task of predicting toxicological endpoints or physical parameters (Enslein and Craig 1978; Cramer 1980). In such problems it has a set of dependent variables, for example, LC50 or vapor pressure, observed on a set of chemicals and wants to explain the behavior of these variables in terms of a set of independent predictors. In a QSAR setting the result is a mathematical equation that provides an estimate of the dependent variable for untested chemicals. This is a classical statistical problem to be solved by multiple regression analysis, linear discriminant analysis, and similar (least-squares) statistical methods. (Enslein et al. 1977).

2.2 Pattern Recognition View

Pattern recognition is a general name attached to a set of methods used to investigate data represented as points in n-dimensional spaces. The field has been described in an enormous literature (Tou and Gonzalez 1974) and many of the techniques have been applied to SAR problems (Kirschner and Kowalski 1979; Stuper, Brugger, and Jurs 1979).

Each compound in the set is represented by a vector of n-molecular structure descriptors,

$$X = (x_1, x_2, \ldots, x_n)$$

where x_j is the value of the j-th descriptor. These molecular descriptors are the independent variables in the statistical analysis. Since a molecular structure can be represented as a vector, mathematical methods of vector analysis can be directly applied to the problem. The mathematics of water analysis bears a strong resemblance to the least-squares methods of statistical analysis. Alternatively the compound may be represented as a point in an n-dimensional space. Then, the value of the point along each of the n orthogonal axes is the value of the descriptor corresponding to that axis. This n-dimensional space, in which each compound is a point, cannot be directly visualized however, it is Euclidean and can be handled mathematically. Thus, the distance between two points in an n-dimensional space is the length of the line joining them; this quantity can be computed using the Pythagorean Theorem generalized to n-dimensional space.

Typically, each structural descriptor has its own origin and scale. In order to allow each variable to contribute equally to the distance measure, they usually are standardized or autoscaled. This maybe done by apply the following equation,

$$x_{ij}' = \frac{x._{ij} - x.}{j}$$

where x_{ij}' is the new standardized variable, x_{ij} is the raw value, x_j is the mean for the j-th descriptor over the data set, and σ_j is the standard deviation for the j-th descriptor. After standardization, each of the descriptors has a mean of zero and a unit standard deviation over the data set.

An assumption invoked in pattern recognition is that two points that are close together in n-dimensional space represent two compounds that are similar with respect to the structural descriptors employed. Just as two compounds that are close together are similar, then a group of points forming a cluster (a region of space where points are densely grouped) defines a class or subclass of compounds.

Methods of pattern recognition investigation include looking for clusters far apart from other clusters, surrounding clusters with surfaces, searching for flat or curved discriminant surfaces that slice the total data space into regions, and constructing models that approach the location of points within one class or another.

Several additional concepts of pattern recognition are as follows. The "training set" refers to a set of compounds that are used to perform the pattern recognition analysis. The actual class to which each member of the training set belongs may or may not be known in advance. If it is known before the calculation is performed, the analysis is considered to be a "supervised" method, if it is unknown, it is an "unsupervised" method. "Validation" refers to the practice of holding a subset of the training set out during the calculations and using the results to classify the members held out. Those compounds held out constitute the "test set." "Prediction" refers to the classification of truly unknown compounds by a model or discriminant.

2.3 Analysis in the Presence of Missing Data

Any method of data analysis discussed in this report could be plagued by missing entries in the data matrix. In particular, it may be that some of the properties of a chemical compound can not be determined or accurately estimated.

One solution to this problem is to discard any compound that has not been completely investigated. It may be undesirable, however, to discard information that was expensive or difficult to collect. In other situations, discarding incompletely studied compounds may mean discarding the whole structure-activity analysis.

Statistical methods for analyzing an incomplete data matrix have been reported. For example, there is an abundance of literature on missing data regression analysis, discriminant analysis, and factor analysis (Beale and Little 1975; Rubin 1976; Hamilton 1975). Most of the available methods are appropriate for data that follow a multivarate normal distribution.

The most popular approach is the E-M algorithm (Dempster, et al. 1976) which estimates the missing elements of the data matrix simultaneously with estimating the parameters of the model. Rubin (Rubin 1974; Ainkins 1980) presented a method for factoring a missing data analysis into smaller, more easily handled subproblems. This factorization also allows one to inspect the pattern of missingness; the pattern may itself be informative.

Most statistical computer packages contain algorithms for treating missing data. There is, however, no standard algorithm and some have been severely criticized (Heiberger 1977).

The primary limitation of a missing data analysis is that it requires fairly strong assumptions about the statistical behavior of the data and about the reasons for missing values. It is important in any particular application, therefore, to know exactly what assumptions form the basis of the method being used and to test the data for the validity of the assumptions.

3.0 METHODS

In the following discussion we differentiate between parametric and nonparametric methods. By parametric methods we mean methods in which the data are assumed to follow a particular probability model, usually a model uniquely represented by mean vectors and covariance matrices. These include the typical statistical methods of regression, discriminants, and principal components analysis, as well as some pattern recognition methods. Caution in using such parametric measures should be exercised since they may not be justified for a particular data set. By nonparametric methods we mean methods wherein the data themselves are used in the development of mappings, clusterings, discriminants, etc. The two categories are, however, not entirely disjoint.

3.1 Regression

In quantitative structure-activity studies, one important goal is to estimate the direction and magnitude of the change in biological response corresponding to a change in molecular properties. In the typical example, the biological response Y (dependent variable) is a linear function of several molecular properties $X_1, X_2, \ldots, X_k$ (independent variables):

$$Y = B_o + B_1 X_1 + B_2 X_2 + \ldots + B_{12} X_{12}$$

The Y and X's of each compound form the data matrix. Usually, the best set of coefficients, B's, are determined by a least-squares, regression analysis (Daniel and Wood 1971). The resulting equation is then used to predict Y's for compounds not used in the regression.

Less often, nonlinear regression analyses are used in attempts to correlate biological response with molecular properties: The model is $Y = g(X_1, \ldots, X_{12}; B_1 \ldots, B_p$, where g is a nonlinear function of the coefficients, $B_1, \ldots, B_p$.

Under well-defined, but not strict, mathematical assumptions, statistical confidence interval and hypothesis testing procedures can be used to assess the validity of the model, to detect outliers, and to select the appropriate dependent variables (Cuthbert and Wood 1980; Draper and Smith 1981).

3.1.1 Data Requirements

Within the data matrix, no single X should be closely approximated by a linear combination of the other X's. The Y variable should normally be measured on a continuum. The biological test system that produces the response Y should be designed so that the Y values are indpendent of each other.

The statistical method of log linear modeling has proven very useful in analyzing data with a categorical response (dependent) variable (Bishop et al. 1975). When there are only two categories, the log linear model amounts to logistic regression, which is often used in epidemiology and bioassay analysis (Cox 1970). Although it has not yet been tried, this approach could be useful in the SAR setting. For a comparison of logistic regression and discriminant analysis see Press and Wilson (1978).

3.1.2 Availability

All standard statistical software packages contain least-squares regression programs (SAS, BMDP, MINITAB, SPSS). The BMDP package, among others, contains a stepwise multiple logistic regression program and a loglinear modelling program.

3.1.3 Limitations

For optimum preductive confidence, it is important that each of the independent variable values used in the regression analysis be spread over the range for which predictions are desired.

It is required that one have data on more compounds than there are independent variables. Particularly when one chooses the best fitting of several posssible regression equations (e.g., by using a step-wise regression program), it is important to have sufficient data to assess the validity of the selected model. Spurious statistical associations are more likely to occur when the number of compounds in the data matrix nearly equals the number of independent variables.

A variety of techniques exists for the selection of subsets of variables in regression and discriminant analysis. The purpose of these subset selection techniques is to find that minimum set of independent variables which "best" or sometimes, most parsimoniously "explains" the endpoint to be modelled (Martin 1978).

These methods include step-up and step-down selection, step-down ridge regression, and "optimal" subset regression (Enslein, Ralston, and Wilf 1977, Dixon 1978).

Caution must be exercised in the use of many predictor variables as this could result in chance correlations being interpreted as meaningful (Topliss and Costello 1972; Kapper et al. 1976; Topliss and Edwards 1977). The scientific plausibility of the regression equation should be ascertained. Although a good fit of the equation to the data may be indicative of a mechanism of action, the good fit is not sufficient evidence that a unifying mechanism is involved (Martin 1978).

It is conventional practice to inspect various plots of the data, residual, and predicted values as part of the evaluation of the regression equation and associated distributional assumption.

It is also preferred to have sufficient data available to validate the selected equation. New statistical methodology such as the jackknife technique and the bootstrap method (Efron 1980) may be particularly useful in QSAR model validation.

Under some circumstances, it may be preferable to estimate the coefficients of the regression equation by methods other than least squares. "Shrinkage methods" such as ridge regression, Stein estimation, and regression on the first few principal components of the independent variables are especially useful if the independent variables are highly correlated (Draper and Van Nostrand 1979). Robust regression methods are useful for identifying outliers and for analyzing dependent variables that violate the usual distributional assumptions (Lanner and Wilkinson 1979; Narula and Wellington, 1979).

3.1.4 Multivariate Considerations

In the event that different responses, denoted by Y_1, Y_2 . . ., Y_m, have been observed for each compound one may wish to analyze simulataneously the relationship between the dependent and the independent variables. There are two basic approaches to analyzing such multivariate response data. The first is canonical correlation, which identifies those orthogonal linear combinations of the dependent variables that best correlate with linear combination of the independent variables (Timm 1975). The second uses knowledge outside of the experiment to collapse the responses into an univariate index (of toxicity or biological activity); the index is then regressed on the independent variables using the regression techniques outlined above. Most statistical computer packages include multivariate regression (canonical correlation) programs as well as the option of transforming the responses variables before regression analysis (SAS, BMDP).

The major limitation of canonical correlation is that it frequently is not always obvious how to interpret the equations in terms of the original data.

56

3.2 Discriminant Analysis for Categorical Endpoints

An alternative to deriving functions that discriminate between groups or classes is to derive separate equations which describe each class. A method based on this approah is SIMCA pattern recognition, which describes each class in a multiclass problem by a disjoint principal components model. A result is that information beyond classification can be obtained (Dunn and Wold 1978). In some cases it is possible, using this method, to derive QSAR's that relate levels of activity to the variables which are found to be relevant to class assignment.

Another advantages to the approach of modelling classes is that in some cases one class, usually nonactives, may not be mathematically definable and separable by a hyperplane from other classes. Rather, separation of the classes in space requires a multidimensional sphere that encloses the active class. This is equivalent to the notion that active compounds are characterized by near-optimum values of many properties whereas an inactive compound may be inactive because any one of these properties is nonoptimum. In such cases principal components modelling can give a classification when discriminant functions methods cannot. The SIMCA program is available on a limited basis from the author.

3.3 Factor Analysis/Principal Components Analysis (FA/PCA)

(For the purposes of the analyses considered here, there is little difference in these methods and for the sake of this discussion, they will be considered identical).

FA/PCA is a method that reduces the dimensionality of a data matrix to a smaller number that still explains most of the variances (Martin 1978; Harman 1967). In the representation of objects (compounds) by n-dimensional data (chemical or pharmacological) each object can be expressed as a vector in n-dimensional space or as a point in n-dimensional space. For the kth compound it is assumed that the mathematical regularity within the data can be approximated by the linear model

$$y_{ik} = m_i + \sum_{a=1}^{A} b_{ia} \, \theta_{ak} + e_{ik}$$

where m_i (i is the variable index 1 to n) is the mean of y_i, a is the number of product terms considered significant to represent the data (a is less than n), b_{ia} is the loading of the ith variable on the 9th component for the component term θ_a and e_{ik} are the residuals for the y_{ik}'s. For $\bar{A}=1$ the model has the form of a straight line, for $\bar{A}=2$ a plane and for $\bar{A} > 3$ the model represents a hyperplane.

The data are usually standardized to zero mean and unit variance. This gives all variables the same metric and reduces the possibility that one variable will dominate or weight the analysis.

If standardized data are analyzed, $m_i = 0$.

If it is assumed that certain variables, i.e., level of biological response in a particular screen, are related to the structural descriptors, the factors may be rotated to relate them to these external variables. This rotation may be usual (eigenvector plots) or it may be quantitative (multiple regression) (Overall and Klett 1972).

If the rotation is quantitative, a QSAR with predictive potential can be the result of the analysis. Therefore one use of this method would be as an alternative to multiple linear regression when regression cannot be used reliably. For example, when the number of descriptors, N, approaches the number of compounds, M, regression should not be used; the ratio M/N should be greater than 4 for regression analysis.

The question of how many A's adequately describe the n-dimensional data matrix is a matter of controversy. For example some consider that the ratio of the number of descriptors, N, to the number of product terms, A, must be 4.(Bender et al. 1973; Ritter and Woodruff 1977). Others consider that only eigenvalues $\geq$1.0 should be considered (Martin 1978); others use that number of eigenvalues that explain a certain fraction of the data (Morrison 1969; Gray 1976).

The FA/PCA method of analysis is available in most statistical packages, i.e., SAS, SPSS, BMDP, etc. These programs are generally available on most machines at major research facilities and are generally used in the batch mode.

3.4 Discriminants

The development of discriminants that separate regions of the n-dimensionsal space from one another is often appropriate for data that is tagged by classes or categories--supervised learning.

Nonparametric discriminant functions are used when the data set to be analyzed does not conform sufficiently to the multivariate normal distribution that is desired for parametric analysis methods. A limitation regarding the data is that one must always limit the number of descriptors per compound to one-third the number of compounds (Bender et al. 1973). One must also ensure thorough testing that the descriptors being used as are not too highly correlated with one another, that is, that multiple collinearities do not exist. (Frank et al. 1965).

A training set of data is presented to the algorithm to develop a surface (usually a flat one) that completely separates the members of one class from another. This presumes a two-class problem, but the methods can generally be extended to multiclass problems as well. The methods used to search for good discriminant surfaces by nonparametric means vary widely but are generally iterative. A heuristic, error-correction feedback method has been reported as used in SAR work--this approach is also known as the linear learning machine or perceptron (Jurs et al. 1979; Chou and Jurs 1979). It works by classifying the members of the training set one by one and correcting its current discriminant surface whenever an error of classification is committed. The algorithm has the highly desirable property that it will converge if possible. Other methods that have been used for the development of linear discriminant functions include least squares methods (Harman 1967) iterative least squares methods (Chou and Jurs 1979) simplex optimization methods (Darvas 1974; Kaderline and Wilkins 1978) and others. These differ in the approach to searching for the separating surface, but not in the overall goal pursued.

The algorithms implementing these discriminant development methods have been reported in the literature. However, these algorithms are not available in commercial software packages.

The results of the development of such discriminants using nonparametric means are the discriminants themselves. These discriminants can be validated by the standard means of randomized testing with a set of randomly selected "test sets" (Weisel and Fasching 1977) Beyond this, it is difficult to ascribe a probability of correctness to the results of a classification by a nonparametric linear discriminant function.

The predicted activity of unknown compounds can be obtained using a discriminant developed as above. The general limittion that operates for all predictive attempts is present here as well - namely, that the unknown compound must be "similar" to those making up the training set from which the discriminate was derived (Frank et al. 1965). This is just to say that interpolations are safer than slight extrapolation which are in turn safer than large extrapolations.

3.5 Cluster Methods

Many clustering methods have been developed (Everitt 1974, Hartigan 1975). These include: hierarchical clustering, minimum spanning tree clustering, centroid clustering, multidimensional scaling, and their variants.

For the sake of brevity we have chosen to limit this discussion to hierarchical clustering.

Hierarchical clustering is used to group compounds according to the similarity of their molecular properties. Primarily two types of questions of a data set are answered by a cluster analysis: "What compounds are similar?" and "How many groups of similar compounds are there?" An attractive potential use is the analysis of data before regression analysis.

The Euclidean distance between compounds is typically measured. That is the square root of the sum of squared distances in each dimension. The directional distances may be in the units of the original data, or they may be standardized by dividing the measure by its standard direction. The distance may also involve higher powers of the dimensional distance. If so, more weight is given to large distances

Several measures may be used for the distance beween multimembered clusters. The centroid method considers it to be that between the center of the two clusters. Alternatively; one could consider it to be the distance between the two closest objects of different clusters or that between the two most distant.

There are no data requirements for cluster analysis (Sokal and Klett 1974). Specifically, one may examine a data set with more predictor variables than compounds. Although there are variations in the available programs, BMDP may be used for this purpose.

Since it is not based on statistical assumptions, the usual statistical validation techniques don't apply. Hence, quantitative predictions are not possible, there is no feeling for the relative importance of the various molecular properties, one has no objective measure for the number of clusters, and it is impossible to predict beyond the data space examined.

3.6 Mapping and Display

For a typical SAR study, each compound will be represented by more than three descriptors. Therefore, the space in which the points representing the compounds are imbedded cannot be viewed directly. However, visualization of patterns is a human capacity that is well documented, so display to a space that could be displayed directly is desirable (Gnanadesikan 1977). Both linear and nonlinear mapping methods have been used (Duda and Hart 1973). Linear methods are exemplified by the Karhunen-Loeve transformation, which is an eigenanalysis of the data. The resulting eigenvectors can be used to linearly transform the original n-dimensional data into any lower dimensional space desired. If the two eigenvectors which have a largest eigenvalues associated with them are used, a two-dimensional set of points results which can be displayed. These two dimensions would be those two orthogonal directions along which the variance of the data are maximized. This transformation has the advantage that a new compound, after it is represented by the same set of molecular structure descriptors as were the compounds used to derive the transformation, can be mapped to the two-dimensional space, that is, predicted. Many software packages contain eigenanalysis routines available that could be used for the purpose of mapping for display.

Nonlinear mapping involves deriving a new two-dimensional space that "best" represents the original n-dimensional space. The most widely used mapping method is that of Sammon (1969) which works by minimizing the following error function:

$$E = \sum_{i<j} \frac{(d_{ij}-d_{ij}^*)^2}{d_{ij}}$$

The goal is to find a two-dimensional space in which all the pairwise interpoint distances are the same in the original n-dimensional space. Obviously, the error cannot be eliminated since it is not possible to exactly represent an n-dimensional space in two dimensions. The error function above is minimized by gradient searching, methods of steepest descant, or other standard mathematical methods. This method of nonlinear mapping and display would be used for direct visualization of a data set. The routines for implementation of nonlinear mapping and display are available in the literature and can be used in conjunction with most other types of pattern recognition analysis. One noteworthy limitation of NLM is its consumption of computer time, especially for large data sets and for large numbers of discriptors.

4.0 APPLICATION TO U.S.E.P.A. PROBLEMS

As shown in sections 2.1 - 3.2.3 above, it is clear that the mathematical methods are available to develop solid SAR models, discriminants, etc. The validity of the models, however, is contingent upon having a good data matrix. Therefore, the application of SAR methodology to problems of direct importance to E.P.A. is dependent on the availability of suitable data and support from the U.S.E.P.A. As the SAR community gains experience in analysis and begins to expand this range of application, some new mathematical methodology may be required.

We consider SAR methods to be informational tools that should be used for internal priority setting within the U.S.E.P.A. SAR methods must be embedded within the overall approach toward decision-making procedures used in the U.S.E.P.A. SAR predictions should not be the sole input to regulatory decisions.

5.0 RECOMMENDATIONS REGARDING RESEARCH OPTIONS

5.1 EPA should take an active role in the organization and accessability of structure-activity data.

5.2 EPA should participate in the support of centers of excellence in SAR for the development of expertise in SAR research, including programs for training investigators.

5.3 EPA should support the generation of high-quality SAR data bases

5.4 EPA should develop a clear set of procedures for the submission of grants in the SAR area, including the personnel directly involved.

5.5 EPA should provide long-term, stable research support for SAR research in addition to short-term, contract support for specific goals.

REFERENCES

Beale, E.M.L. and Little, R.J.A. 1975. "Missing Values in Multivanite Analysis" Jour. Royal. Stat. Soc. B 37: 129-145.

Bender, L.F., Shepard, H.D., and Kowalski, B.R. 1973. Alternative representation of mas spectra for the spectral identification problem. Anal. Chem. 45: 617-618.

Bishop, Y.M.M, Bienberg, S.E., and Holland, P.W. 1975. *Discrete Multivariate Analysis: Theorgy and Practice*, The MIT Press

Chon, J.T. and P.C. Jurs, 1979. Computer assisted structure - activity studies of chemical carcinogens. An N-mitroso compound data set. J. Med. Chem. 22: 792-797.

Cox, D.R. 1970. *The Analysis of Binary Data*, Barnes and Noble

Daniel, C.; Wood, F.S. 1971. Fitting equations to data. New York: Wiley.

Cuthbert, Daniel and Wood, Fred S. 1980. *Fitting Equations to Data: Computer Analysis of Multifactor Data* 2nd edition, John Wiley & Sons

Darvas, F. 1974. Application of the sequential simplex method in designing drug analogs. J. Med Chem. 17: 799-804.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1976). "Maximum Likelihood from Incomplete Data via the E.M. Algorithm Research Report S-38, NS-320, Dept. of Statistics, Harvard University.

Draper, N.R. and Van Nostrand, R.C. 1979. "Ridge Regression and

James-Stein Estimation: Review and Comments", *Technometrics* 21: 451-466

Draper, N.R. and Smith, H. 1981. *Applied Regression Analysis, 2nd edition,* John Wiley & Sons.

Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis. Wiley, New York, 1973

Dunn, W.J., III and Wold, S. 1978. A structure-carcinogenitity study of 4-nitroquinoline toxide using the SIMCA method of pattern recognition. J. Med. Chem 21: 1001-1007.

Efron, Bradley (1980) The Jackknife, The Bootstrap, and other Resampling Plans, Tech. Report # 63, Div. of Biostatistics, Stanford University.

Everitt, B., Cluster Analysis, Wiley, New York, 1974

Frank, R.E. Massy, W.E. and Morrison, D.G. 1965. Bias in multiple discriminant analysis. J. Marketing Research 2: 250-257.

Gnanadesikan, R., Methods for Statistical Data Analysis of Multivariate Observations. Wiley, New York, 1977.

Gray, N.A.B. 1976. Constraints on "learning machine" classification methods. Anal. Chem. 48: 2265-2268.

Hamilton, M.A. 1975. "Regression Analysis when there are Missing Observations--Survey and Bibliography" Statistical Center Technical Report #1-3-75, Montana State University.

Harman, H.H. 1967. Modern factor analysis, 2nd ed. Chicago: U. of

Chicago Pr.

Hartigan, J.A., Clustering Algorithms, Wiley, New York, 1975.

Heiberger, R.M. (1977) "Regression with the Pairwise-Present Covariance Matrix: A Dangerous Practice" Technical Report #19, Dept. of Statistics, Univ. of Pennsylvania.

Hinkins, S. (1980) "RFACTOR - A Program to Create Rubin's Factorization when there are Incomplete Multivariate Data". Am. Stat 34: 182-3.

Jurs, P.C., Chon, J.I., and Yuan, M. 1979. Computer-assisted structure-activity studies of chemical carcinogens. A heterogenous data set. J. Med. Chen. 11: 476-483.

Kaberline, S.L. and Wilkins, C.L. 1978. Evaluation of the super-modified simplex for use in chemical pattern recognition. Anal. Chimica Acta 103: 417-428.

Kirschner, G.L. and Kowalski, B.R., The Application of Pattern Recognition to Drug Design, in Drug Design, Vol. VIII, Ariens, E.J., (Ed.), Academic Press, New York, 1979.

Kowalski, B.R. and Bender, C.F. 1974. The application of pattern recognition to screening prospective anticancer drugs. Andenocarcinoma 755 biological activity test. J. Amer. Chem. Soc. 93: 916-918.

Kupper, L.L. Stewart, J.R., and Williams, K.A. (1976). "A Note on Controlling Significance Levels in Stepwise Regression".

Am. Jour. Epid. 103: 13-15.

Launer, R.L. and Wilkinson, G.N. (1979). Robustness in Statistics, Academic Press.

Martin, Y.C. 1978. Quantitative drug design. New York: Dekker; pp 167-232.Morrison, D.G. 1969. On the interpretation of discriminant analysis. J. Marketing Research 6: 156-163.

Narula, S.C. and Wellington, John F. (1979). "Multiple Linear Regression: A Review of Two Alternatives to Least Squares Criterion". Res. Report #37-79-p. 10, Operations Research and Statistics, Rensselaer Polytechnic Institute, 96 references.

Overall, J.F., Klett, C.J. (1972). Applied Multivariate Analysis. New York: McGraw.

Press, S.J. and Wilson, S. (1978). "Choosing Between Logistic Regression and Discriminant Analysis "Jour. Am. Stat. Assoc. 73: 699-705.

Ritter, G.L. and Woodruff, N.L. 1977. Dimensionality and the number of features in "learning machines" classification methods. Anal. Chem. 49: 2116-2118.

Rubin, D. (1974). "Characterizing the Estimation of Parameters in Incomplete-Data Problems". Jour. Am. Stat. Assoc. 69: 467-474.

Rubin, D.B. (1976). "Comparing Regressions When Some Predictor Values are Missing". Techno. 18: 201-205.

Sammon, J.W. 1969. A Nonlinear Mapping for Data Structure Analysis IEEE Trans. Comput., C18, 401.

Stuper, A.J., Brugger, W.E., Jurs, P.C., 1979. Computer Assisted Studies of Chemical Structure and Biological Function Wiley-Interscience, New York.

Sokal, R.R., Kleth, C.J. 1974. Classification: Purposes, Principles, Progress, Prospects. *Science 185*. 1115-1123.

Timm, Neil H. (1975). *Multivariate Analysis with Application in Education and Psychology*, Brooks/Cole (esp. Chapter 4 "Multivariate Regression and Canonical Correlation Analysis").

Topliss, John G. and Costello, Robert J. (1972). "Chance Correlations in Structure - Activity Studies Using Multiple Regression Analysis". *Jour. Med. Chem*. 15: 1066-1068.

Topliss, J.G. and Edwards, R.P. 1979. Chance factors in studies of quantitative structure-activity relationships. J. Med. Chem. 22: 1238-1244.

Tou, J.T. and Gonzalez, R.C., 1974. Pattern Recognition Principles Addition-Wesley, Reading, Mass.

Weisel, C.P. and Fasching, J.L., 1977. Deceptive "correct" separation by the linear learning machine. Anal. Chem. 49: 2114-2116.

GROUP II

TABLE OF CONTENTS

GROUP II - EVALUATION OF MOLECULAR DESCRIPTORS

	Page
Background and Summary	1
Molecular Descriptor 1: Partition Coefficient	3
Evaluation of State of the Art	3
Research Needs	6
Recommendations	6
Molecular Descriptor 2: Solubility	8
Evaluation of the State of the Art	8
Research Needs	9
Recommendations	10
Molecular Descriptor 3: Molar Refractivity	11
Evaluation of State of the Art	11
Research Needs	12
Recommendations	13
Molecular Descriptor 4: Hydrophilicity	14
Evaluation of State of the Art	14
Requirements and Limitations	14
Research Needs	14
Recommendations	15
Molecular Descriptor 5: B,C-Parameter	17
Research Needs	18
Recommendations	19
Molecular Descriptor 6: X-T Parameters	20
Evaluation of State of the Art	20
Research Needs	20
Recommendations	22
Molecular Descriptor 7: Degree of Ionization	24
Evaluation of State of the Art	24
Recommendations	25
Molecular Descriptor 8: Molecular Connectivity	26
Evaluation of State of the Art	26
Research Needs	27
Recommendations	28

TABLE OF CONTENTS (cont.)

Page

Molecular Descriptor 9: Substructural Key Systems 30
Evaluation of State of the Art 30
Research Needs 30
Recommendations 30

Molecular Descriptor 10: Molecular Volumes and Areas 31
Evaluation of State of the Art 31
Research Needs 31
Recommendations 32

Molecular Descriptor 11: Descriptor Reduction 33
Evaluation of State of the Art 33
Research Needs 34
Recommendations 35

Molecular Descriptor 12: Theoretical Chemical Descriptors 36
Evaluation of State of the Art 36
Research Needs 36
Recommendations 39

References 57

LIST OF FIGURES

Figure Number — Page

1 Suggested Procedure for Optimal Collaboration 37
2 Capabilities of Theoretical Chemistry Relevant to the Selection of Molecular Descriptors (Steps 3 and 4) 38
3 Characterization of Nonbonding Interactions (Complex Formation) 39
4 Capabilities of Theoretical Chemistry Relevant to Calculation of Molecular Descriptors: Screening (Step 5) 40
5 Specific Methods 41
6 Classes of Chemical Carcinogens 43
7 Untitled 44
8 Structure of BP-Guanine Adduct 45
9 Untitled 47
10 Proposed Mechanisms for Oxidation of PAH by Cytochrome P-450's 48
11 Methods Used 49
12
13
14
15 Three Alternative Hypotheses for PAH C Carcinogenicity 52
16
17
18 Proposed Metabolic Pathways for the Anesthetic Enfluane 53
19 Quantum Chemical Studies of Anesthetic Metabolism 54
20
21
22 Conclusions 55
23 Conclusions (Mechanistic) 56

LIST OF TABLES

Table Number — Page

I - IX
X Calculated Energy of Carbocation Formation and Observed Carcinogenic Potency for 44 PAH 50

GROUP II - EVALUATION OF MOLECULAR DESCRIPTORS

Background and Summary

The molecular descriptors used in Quantitative Structure Activity Relationships have, historically, included many types not calculated directly from structure. The first such descriptor was, in fact, a parameter based on a physical property: the partition ratio of the solute between in immiscible lipid phase and water. It was used near the turn of this century by Meyer and Overton (Meyer 1899, Overton 1901) to rationalize narcotic action on tadpoles. Next on the scene came the 'extra-thermodynamic' electronic parameters pioneered by Hammett (Hammett 1935, 1970) followed by steric parameters introduced by Taft (Taft 1956). Although both types have as their primary data source rate and equilibrium constants, some encouraging progress has been made in substituting for Taft's E_s a set of steric constants derived from van der Waal's radii and thus available directly from structure (Verlopp et. al., 1976, Charton 1975).

Topological descriptors, such as molecular connectivity (Randic 1975, Kier and Hall 1976), are structurally derived but would be of limited utility without incorporating in the topology of each molecule some higher level of information about the nature of any heteroatoms it may contain. Valence factors derived from the number of valence and non-valence electrons or else from the familiar field, resonance and molar refractivity parameters can be added to accomplish this end (Yalkowsky 1980). Of course the ultimate in understanding would be achieved when atomic structure is known in such detail that essentially all molecular properties, both chemical and physical, could be predicted prior to synthesis and testing. Quantum chemistry, based on the Schroedinger equation has provided at least the beginnings of an approach to this theoretical ideal. Various semi-empirical approaches have yielded useful information on conformation, ionization potentials, energy levels of activated intermediates, etc. (Kier, 1971).

From the foregoing it should be apparent that the subject of molecular descriptors can be categorized in several ways based on several perspectives. A historical development of descriptors also traces a trend from the more pragmatic to the more theoretical; from a dependence upon measurements of physical properties to a greater (or total) reliance upon involved computer calculation. In the following sections the descriptors will be discussed in their approximate chronological order of appearance: a) those derived from physical properties, e.g., partition coefficient, molar refractivity, hydrophilicity, solubility; b) those based on reaction rate or equilibria and used in linear free energy relationships, e.g., Hammett sigma and Taft E_s; and c) those calculated from structure, e.g., connectivity, upsilon, molecular volume and cavity area, structural 'keys', and quantum chemical indices.

MOLECULAR DESCRIPTOR 1: PARTITION COEFFICIENT By A. Leo

Evaluation of State-of-the-Art

The historical development of this 'granddaddy' of biological QSAR parameters, the rationale of using octanol/water as the reference system and especially the use of log $P_{o/w}$ to make it compatible with other free-energy-based parameters is adequately covered in the literature (Hansch 1979, Leo 1971). This development also serves to delineate its scope and limitations. The important points to stress when considering its application to environmental QSAR are these: a) a totally non-polar lipid phase would not be an appropriate model, because it completely excludes ions and other very polar solutes, and furthermore it would need finely-tuned hydrogen bonding parameters to represent binding or membrane transport in the environmental milieu. Log $P_{o/w}$ has been shown to be related to solute binding to surfaces as widely divergent as carbon, nylon polymer and protein (Hansch and Helmer 1968, Dunn and Hansch 1974). However, where differences in the hydrophobicity of small anions may be important, log $P_{o/w}$ may lack sufficient sensitivity (Leo 1971), and log $P_{oil/air}$ relates better to lung/air transport. b) Because it is not a simple parameter but a measure of the relative solvation energy in two mutually saturated phases, log $P_{o/w}$ discriminates between very slight differences in the structure of polar groups, a fact very much in evidence in a table of fragment values used in calculation methods described below (Hasch 1979). c) One must realize that the ratio (octanol solvation energy) (aqueous solvation energy) may differ greatly from $P_{o/w}$, especially for the very lipophilic solutes of paramount interest to the U.S. EPA. This apparently arises from the 2M water in the saturated octanol phase and the 10^{-3}M octanol in the saturated water phase--the former fortuitously modeling the polar character of many proteins and bio-lipids and the latter modeling the low level of lipid material in even 'pure' lakes and rivers. While more data characterizing this difference between $P_{o/w}$ and the ratio of solubilities would certainly be desirable, that now in hand indicates that bioaccumulation is predicted more reliably by the former (Branson 1976).

Because of their long established use in bio-molecular design, there exists a useful data base of log P values (Leo 1981), but except for a sizable number of values for pesticides, most of the partition coefficients needed by the U.S. EPA have not been measured. Even if the present modest level of support for P.C. measurements were to greatly increase, there would still be a backlog of unmeasured solutes for many years to come. The need for a reliable method of calculating these values based on structure has long been recognized. Because of the problems associated with the mutual saturation of phases referred to above, an empirical approach to log P calculation (Rekker 1977 , Hansch 1979) was necessary if EPA's current problems were to be addressed.

The limitations of obtaining suitable measured partition coefficients are these: The 'standard' shake-flask procedure is the most dependable but very time-consuming if the standard error of measurement is determined and minimized; more rapid HPLC procedures can yield values closely related to log $P_{o/w}$ but certain solute structures need correction factors to account for unique binding to the support (Unger 1979). These corrections may not all be apparent from structure. On the other hand, for very lipophilic solutes (log $P_{o/w} > 5.5$) where analytical procedures place severe limitations on the precision of the shake-flask method, HPLC may be the procedure of choice (Veith).

The log $P_{o/w}$ calculation method being developed by EPA is a variation of one proposed by Rekker (Rekker 1977), modified by Leo (Hansch 1979) and adapted to a computer algorithm by Chou et. al., (Yalkowsky 1980). The CLOGP program accepts structural input hand drawn on a CRT or, with an interfaced WISCT program (Leo 1981), as Wiswesser Line Notation. Input via CAS connection table is an option under development. The output lists all fragment constants and correction factors and calculates log $P_{o/w}$ as the neutral structure. Appropriate messages instruct the user about corrections for the ionic forms of many acids and amines where these can be reliably estimated.

At this early developmental stage, there appear to be two serious limitations to CLOGP. The present fragment identification procedure has a practical upper limit of about 250 fragments, which has already been reached. Secondly, the effect upon a hydrogen bonding fragment by one other electronegative fragment has been allowed for, but multiple substitution of this type can lead to serious deviations.

Research Needs

Log $P_{o/w}$ has been shown by several investigators to be related to bioaccumulation (Neeley 1974, Veith   ). Its relationship to binding to the organic component of soil, K_{oc}, (Kenaga 1980) makes it valuable in transport models. Because it models non-specific hydrophobic interactions, it can be a useful predictor of certain types of toxicity, such as narcosis, hemolysis and necrosis, but one should not expect that a linear relationship of the sort which is found with lower homologs will be maintained indefinitely.

Recommendations

Listed in order of priority, the recommendations to the U.S. EPA for initial support are: comprehensive literature search and data base maintenance and design; measurement by HPLC of very lipid solutes; and further testing of the structural limitations to the relationship of log $P_{o/w}$ to log K_{oc} and log k_1/k_2 (bioaccumulation). Continued support is recommended for the measurement by standard shake-flask procedure of solutes whose structures will add new fragment values for CLOGP and for improvements in the CLOGP algorithm itself.

MOLECULAR DESCRIPTOR 2: SOLUBILITY by S. Yalkowsky

Evaluation of State-of-the-Art

It has been shown that the aqueous solubility of liquid and crystalline nonelectrolyte solutes can be estimated if it is dividied into two components (Yalkowsky 1980): the ideal solubility X_i and the solubility of the liquid form of the solute X_w. The latter term represents either the oil or the hypothetical supercooled liquid. The total aqueous solubility, X_w, is then given by: log X_w = log X_i + log X^L_w (Eq. 1).

The relationship between solubility of liquids and the octanol/water partition coefficient, $P_{o/w}$, proposed by Hansch et al., (Hansch 1968) has been modified and extended. A single equation works very will for liquid nonelectrolytes:

$\log X^L_w = -\log P_{o/w} + \text{const.}$ (Eq. 2).

By considering the crystal interactions which reduce the ideal component of solubility it is possible to extend the above relationship to crystalline nonelectrolytes. The reduction of the ideal solubility, which is also called the fugacity ratio, is given by:

$$\log X_i = \frac{-\Delta S_f(Tm - T)}{2.303\ RT} \quad \text{(Eq. 3); at room temperature:}$$

$$\log X_i = \frac{-\Delta S_f(MP - 25)}{1364} \quad \text{(Eq. 4)}$$

For liquids, MP is replaced by 25°C and $\log X_i$ becomes zero. Fortunately ΔS_f can be predicted with reasonable reliability: for rigid molecules, $\Delta S_f \approx 13.5$; for flexible molecules, $\Delta S_f \approx 13.5 + 2.5(n-5)$.

The incorporation of these values into the ideal solubility equation gives: $\log X_i \approx -0.01(MP-25)$ for rigid molecules (Eq. 5) and $\log X_i \approx -[0.01 + 0.0018(n-5)](MP-25)$ for flexible ones (Eq. 6).

The aqueous solubility of a crystalline substance is estimated by inserting Eq. 2 and either Eq. 5 or Eq. 6 into Eq. 1. This method has been shown to work for nonelectrolytes which are either rigid or largely flexible. It has not yet been verified for weak electrolytes and for molecules which are partially flexible. It requires an experimental melting point for solids and the measurement or calculation of $\log P_{o/w}$ and ΔS_f.

Research Needs

A knowledge of solubility is essential for understanding and predicting the fate of compounds in the aqueous environment. Since solubility limits the amount of material that can be taken up by the organism, it becomes a major determinant of biological activity. Solubility controls the dissolution rate of solid or oily residues, and so it can be used to estimate the time course of pollution resulting from a single spill. It can also be used to calculate the maximum amount of compound that can be carried away by a river. It will be seen that, because of these factors, solubility can be a limiting parameter in the bioaccumulation of organic materials.

Recommendations

It is recommended that the U.S. EPA financially support a coordinated effort to obtain reliable aqueous solubility data through literature search and measurement. It should also support research in correlating solubility with other physical and biological properties.

MOLECULAR DESCRIPTOR 3: MOLAR REFRACTIVITYby A. Leo

Evaluation of State-of-the-Art

Molar refractivity (MR) for liquids can be calculated according to the Lorentz-Lorentz equation:

$$MR = \frac{\eta^2 - 1}{\eta^2 + 2} \cdot \frac{MW}{d}$$

where η is the index of refraction using sodium light at 20°C; MW is the molecular weight and d is density measured at 20°C. The physical properties needed, η and d, are easily measured using simple equipment, and when the additive-constitutive nature of MR was established (Vogel 1948), this parameter was used as a confirmation of structure along with elementary analysis.

In their earliest efforts to devise an accurate calculation based on structure, Vogel's group used atom values. Later they switched to bond values, and since there were more ways to combine the variance between the calculations (structure vs. L-L). Quite a number of functional groups and structural features were included in Vogel's studies, but no comprehensive table of values was published nor were all the ring and conjugation corrections assembled in one report. Furthermore, no statistical justification of the correction factors was presented.

Preliminary investigations with a data base of 900 structures incorporating a much wider variety of atoms and functional groups than Vogel used (Leo 1981) indicates that a 'hybrid' scheme employing atomic values plus corrections for 'unusual' bonds can be devised to use the minimum number of structural parameters. It is well adapted to computer calculation and promises to produce MR values with a deviation of less than 1% from the Lorentz-Lorentz calculation. This is seen to be well within the range needed for predictive purposes for hazard evaluation.

Molar refractivity of low melting solids can be calculated via the Lorentz-Lorentz equation, and for congeneric series whose lower members are liquid, these show no marked deviation from the values calculated from structure. Refractive indices and density data of sufficient precision is available from a number of sources--the Aldrich catalogue and the Russian organic chemistry journals are two rich sources.

Research Needs

MR primarily relates to molar volume, but it also has a component of polarizability which has been shown to be useful in modeling dipole-dipole interactions or binding due to London dispersion forces (Hansch 1979). It also appears uniquely useful in correlating solute binding to (or inhibition of) isolated enzymes, a factor which certainly should be considered in toxicity evaluations.

Recommendations

Molar refractivity has not yet been established as an important parameter for environmental QSAR, but very little effort has been made as yet to do so, except to calibrate halogen valences in the molecular connectivity approach (Kier 1976). Therefore a high priority cannot be recommended for EPA support for MR development, because some work in this area will undoubtedly be carried on without it. Nevertheless, a very modest level of financial a support would assure: 1) that the data base of structures for which Lorentz-Lorentz calcuation could be made would be expanded to several thousand and would include structures of particular interest to EPA; 2) that the computer calculation program would be compatible with EPA's facilities and be available in the shortest possible time; and 3) that investigations are made into the separate contributions of the components of molar volume and polarizability to determine if the combination in use at present is optimal.

MOLECULAR DESCRIPTOR 4: HYDROPHILICITY By J. Hine

Evaluation of State-of-the-Art

The hydrophilicity of a molecule, γ, is defined as the distribution coefficient of the compound between dilute aqueous solution and the gas phase:

$$\gamma = \frac{\text{conc.}_{Aq}}{\text{conc.}_{gas}}$$

A number of gamma values have geen tabulated (Hine and Mookerjee 1975). Values of log γ have been expressed as additive and constitutive functions of structural components of the molecules involved.

Requirements and Limitations

In order to determine the hydrophilicity of a molecule, its vapor pressure should be large enough to measure. If the vapor pressure is measurable, there will essentially always be enough water solubility, but occassinally compounds will react too rapidly with water and the hyrophilicity will not be measureable. The correlation in terms of molecular structure is good for monofunctional compounds but is sometimes poor for polyfunctional compounds.

Research Needs

Hydrophilicities are directly relevant to the following types of EPA questions:

1) What is the tendency for a compound in aqueous solution to go into the vapor phase; i.e., from a river, lake, or ocean to the air; or from the wet surface of lung into the air, or the reverse?
2) What is the tendency for a compound to go from the vapor phase into aqueous solution; i.e., how well will rain wash a pollutant from the air, or how much inhaled pollutant will go into the moist surface of a lung?
3) The correlations in terms of molecular structure permit one to estimate the vapor pressure of a compound from its water solubility, or to estimate its water solubility if its vapor pressure is known.

Recommendations

EPA should support a comprehensive literature search and measurements where necessary to determine more hydrophilicity values and thus increase the representation of certain structural features that are common to industrial chemicals and to biologically active compounds. This should permit a substantially improved understanding of the proximity effects--the deviations from "ideal" behavior shown by polyfunctional compounds.

Reference

Hine, J. and Mookerjee, P.K., 1975. Intrinsic Hydrophilicities of Organic Compounds. Journal of Organic Chemistry 40:292-8

MOLECULAR DESCRIPTOR 5: B,C-PARAMETER, by R. Cramer

Evaluation of State-of-the-Art

It has recently been shown (Cramer 1980) that many of those molecular properties which are dependent upon non-covalent, orientation-averaged molecular interactions can be interrelated to 95% by a two parameter equation ("BC") and to 99% by a five parameter equation ("BCDEF"). Properties which clearly can be interrelated in this way include partition coefficient ($P_{o/w}$) and molar refractivity (Hansch 1979), hydrophilicity (Hine 1975) and molecular (molar) volume (Perlman   ). Subsequent investigations indicate that the degrees of interrelation indicated in this reference appear to be maintained among much larger and/or more polar compounds.

BCDEF parameter can be obtained for unknown compounds in one of two ways:

1) The preferred way is by fitting some known physical properties of compounds. A minimal set of compounds of interest to EPA might be: measured partition coefficient, molecular weight, melting point, calculated molar refractivity, Bondi molar volume, and magnetic susceptibility (Pascal rules of calculation).
2) Less desirable is a calculation direct from structure using an additive-constitutive model proposed by Cramer (Cramer 1980).

93

Probably a hybrid of the above two approaches may also be useful, but it is recommended that an experimental partition coefficient ($P_{o/w}$) be considered as the minimum in terms of measured physical properties.

Research Needs

A consensus exists that today's QSAR methodologies are more likely to help in the definition of the concentration of pollutant at some locale, rather than the toxicological implications of that concentration. In this connection it seems that the BCDEF interrelationship might help predict the behavior of unknown compounds in either of two ways:

1) If kinetic or physical models governing this fate of molecules in the environment can be established, then the BCDEF interrelationship might define the parameters which control the behavior of an unknown in this model; e.g., binding to clay, vaporization from water, partitioning into fatty substances.

2) Failing the existance of such explicit kinetic models, it may be possible to proceed in this fashion: If one can define and measure the concentration at some location, 'B', as being at the threshold of environmental risk and as a function of the chemical leaked at some other point, 'A', and carry out this experiment for twenty other compounds

chosen from the list given in J. Am. Chem. Soc., 102, 183 (Cramer 1980), it is reasonable to hope that the BCDEF parameters will model this process, independent of its mechanistic detail.

If neither condition (1) or (2) define end points of interest to EPA, then it is difficult to see how any QSAR approach can help, since all QSAR methodologies rely on fitting parameters to end points.

Recommendations

The pioneering work with the BC-Parameter was undertaken at a drug company to clearly separate non-specific biological effects from those which might be optimized to yield a marketable product. It has succeeded to the extent planned, and it is unlikely that this study will be continued privately. Since the separation of non-specific toxicity from unique toxicity is of high priority to EPA, and the BC-Parameter promises to be very helpful in this application, a rather high priority should be assigned to its support. It will be noted that in the minimal set of properties necessary for BC-Parameter calculation, measurements of some are already receiving support. Support for others should be considered as well as support for their inclusion in an accessible data base.

MOLECULAR DESCRIPTOR 6: X-T PARAMETERS, BY M. Charton

Evaluation of State-of-the-Art

'Extra-thermodynamic' (X-T) parameters of the type introduced by Hammett (Hammett 1970) and developed by Taft (Taft 1956) and Hansch (Hansch 1979) are the basis for the linear free energy (LFER) approah to QSAR. In the 'primary' set of X-T parameters, those measuring electronic effects were derived from the pKas of substituted benzoic acids and those measuring steric effects from the rates of ester hydrolysis. In the course of time, 'secondary' sets of X-T parameters were developed which offered at least two distinct advantages: the field and reasonance electronic effects could be more clearly separated (e.g., into σ_I and σ_D) giving a clearer insight into mechanisms, and the steric constants could be obtained from van der Walls radii and calculations made for substituents for which hydrolysis rates were either impossible or impractical to obtain (μ) (Charton 1980). These 'secondary' parameters (all no longer thermodynamically derived) are scaled to be used interchangably with the 'primary' and have greatly extended the applicability of the LFER approach.

Research Needs

96

The number of available σ_I and σ_D constants is insufficient for EPA's purposes, and in many cases the values known are of doubtful quality. The cost of determining experimentally all of the necessary constants in terms of both time and money is exhorbitant. The most reasonable solution to the problem is the development of methods for the estimation of these X-T parameters.

It was shown (Charton 1981) that for some set of substituents WZ, where W is a group such as (-P=O) and Z is an electron acceptor, both the σ_I and σ_D values could be obtained by regression analysis on a large, well characterized data set. Unfortunately, such data sets are generally unavailable. For example where W is (-P=O) there are no values for an electron acceptor such as ($-CF_3$). Similarly for the estimation of σ_I and σ_R of ZNH groups the corresponding measured values for MeONH and MeSNH would be very helpful. It is extremely important that the estimation equations be based on the widest possible range of substituent Z, as only values found by interpolation can be regarded as reliable estimates.

A method of considerable importance in estimating electrical effect constants is based on the observation that substituent constants for groups of the type, $W-AK_n$, (where W is constant and connected to a alkyl chain of variable length) exhibit constant values of σ_I and σ_D. The experimental determination of three

97

of four sigma values for each case would make possible the estimation of a very large number of values for related structures.

The following equation represents another method of estimating σ_I values:

$$\sigma_I,WZ_n = a_1 \chi_w + a_2 n + a_o$$

where χ_w is the Allred-Rochow electronegativity (Allred and Rochow (1958) of W and 'n' is the number of constant Z groups attached to W. The use of this equation permits a crosscheck with estimates of σ_I from other equations, as well as providing estimates for some groups which cannot be made in any other way.

Recommendations

The application of LFER QSAR methods is now accepted in the field of drug and pesticide design. Its success is due in part to the method's simplicity and its economy of time and money. Much of this has been accomplished using X-T parameters of the types discussed in this section, but, on the other hand, many other efforts were thewarted for lack of values. The task of hazard assessment facing OTS encompasses a much broader class of chemicals and thus requires a much more extensive data base than that barely adequate for the drug and agrochemical industries. It is recommended that at least medium priority be given to support of X-T parameter measurement (e.g., there is a real need for more σ_R^+ and σ_R^- values), and for a critical compilation of those reported in the literature with special attention to estimates of error in these values.

MOLECULAR DESCRIPTOR 7: DEGREE OF IONIZATION

Evaluation of State-of-the-Art

It would seem that the degree of ionization, derivable from the pKa, which is itself an X-T parameter, ought to be included in the previous section, and, in fact, ought to be adequately covered when the appropriate sigma values are used. This has not proved to be the case in actual practise, as evidenced by a statement by Corwin Hansch in the foreword of a recent book on the physical chemical properties of drugs (Yalkowsky 1980): "Yvonne Martin tackles the most difficult and crucial problem of forming . . . QSAR with sets of congeners ionized to different degrees, . . . Since many drugs and biological important molecules are more or less ionized at physiological (and environmental) pH, the importance of Martin's pioneering work is evident to all of us involved in QSAR . . . Perrin, using additivity principles and Hammett's basic equation, illustrates how pK_a values for a wide range of complex molecules can be calculated."

(Dr. Martin took part in the activities of Group I, but asked that this short summary be included in this section's report. The session chairmen concur.)

The method of pK_a calculation presented by Perrin (Yalkowsky 1980) provides a good foundation, but there is room for improvement, for the error is ± 0.4 in many cases. This effort would be aided materially if there were a comprehensive collection of pK_a's in computer-readable form.

Recommendations

Suggested areas for EPA support are:

1) Establishment of a data base, searchable down to connection tables, of existing pK_a data.
2) Research on methods to calculate pK_a' from structure at least comparable to those now available for conformation and log $P_{o/w}$.
3) Research on how to treat ionization within the context of QSAR. Some drug research efforts in this area may have implications for EPA problems; e.g., the efforts to change urine pH to help the body eliminate an overdose of a drug (acid pH for bases such as amphetamines and alkaline pH for acids) may suggest some "ion-trapping" procedures of use to deal with environmental hazards.

MOLECULAR DESCRIPTOR 8: MOLECULAR CONNECTIVITY By L. Hall

Evaluation of State-of-the-Art

Molecular connectivity is a method for generating molecular descriptors directly from molecular structure (Kier and Hall, 1976). The method generates a set of connectivity indexes which are a quantitation of significant structural information useful for structure-activity relationships. A recent chapter (Kier 1980) summarizes its development, discusses many applications, gives some detailed examples of how structure interpretations are made from SAR equations, and gives a nearly complete bibliography.

Molecular connectivity has been used in a wide variety of SAR studies. Water solubility, heat of vaporization, boiling point and octanol/water partition coefficients are some of the physical chemical properties for whch model studies have been done (Kier and Hall 1980; Hall, Kier, and Murray 1974; Murray, Hall and Kier 1975). The specific input for molecular connectivity calculation consists of a connection table and atom identifiers (atom number, number of valence electrons and number of hydrogen atoms bonded to the given atom). From this input a large set of connectivity indexes may be computed. Recent work at ERL-D has developed computer algorithms which optimize the computation so that computation time for large molecules is very reasonable.

Current work focuses on further development of the hetero atom parameters. The atom parameters used in the calculations are related to atomic numbers and number of bonded hydrogens as:

$\delta v = (Z^v - h) / (Z - Z^v_{-1})$. These parameters have been widely used but need to be evaluated further against several data sets for certain atoms of interest to the EPA mission, especially phosphorous and sulfur. Further work is being done on the relation of the atom parameters to electronegativity and van der Waal's volume (Kier and Hall 1980), and may even provide the basis for the estimation of Hammett type substitutent constants (Hall 1980).

Research Needs

Molecular connectivity indexes have been used in three areas which are directly related to EPA needs:

1) Physical properties such as solubility, volatility and $P_{o/w}$ are necessary in environmental fate models, and connectivity has been useful in estimating such properites directly from structure.

2) In references given above there are some toxicity correlations which have used connectivity. ERL-D has supplied LC50 data on the fathead minnow from which the following correlation equations were derived:

a) For 22 phenols:

$$pLC50 = -.22\ {}^{1}\chi - .81\ {}^{3}\chi_{P}^{v} - 1.7; \qquad r = .91;\ s = .29$$

b) For 32 alcohols, ethers and ketones:

$$pLC50 = -1.0\ {}^{1}\chi^{v} + .28; \qquad r = .94;\ s = .42$$

c) For 9 halocarbons:

$$pLC50 = -.5\,{}^{o}\chi^{v} - .62 \qquad r = .97;\ s = .22$$

d) For all 63:

$$pLC50 = -.68\,{}^{o}\chi^{v} + .52; \qquad r = .89;\ s = .59$$

$$pLC50 = -.52\,{}^{o}\chi^{v} - .29\ {}^{3}\chi_{p} + .23;\ r = .91;\ s = .53$$

More studies of this type can provide a base line for general narcosis and a means of identifying suspected highly toxic materials.

3) The structural information contained in connectivity indexes may be used with pattern recognition techniques to help classify highly toxic substances.

Recommendations

EPA support is recommended for the following:

a) Connectivity indexes can be made more readily accessible to the QSAR community on a broad basis. The computer programs which are already available can be distributed or access encouraged to a central data base, or a combination of the two.

selected from the literature or obtained from experiment where necessary, and model studies performed. Predictive ability should then be tested by considering data not included in the original data set.

c) Classification schemes should be investigated which are intended to give broad indications of toxicity or level of effect. Predictive equations can be developed to estimate magnitude of effect.

MOLECULAR DESCRIPTOR 9: SUBSTRUCTURAL KEY SYSTEMS, By K. Enslein

Evaluation of State-of-the-Art

Various substructural key systems have been used in SAR. These include CIDS and CROSSBOW. The latter is based on Wiswesser Line Notation (WLN) and was originally developed for ICI in England. The original 148 CROSSBOW keys have been augmented by further descriptors, including those substructures considered as suspicious of carcinogencitity by FOA Bureau of Foods. The present list thus consists of about 350 keys.

There is presently under development (Enslein) another set of keys which will, hopefully, be more meaningful chemically than the WLN-based keys. These keys are based on a system developed some time ago at Smith, Kline and French. With appropriate support it would be possible to input these keys into a graphics system that would also be compatible with the Chemical Abstracts connection tables.

Research Needs

Encouraging results (Enslein) have been obtained using these keys in analyzing rodent toxicity and carcinogenicity data. The chances are good, therefore, that they would prove effective in studying a wide variety of environmental hazards.

Recommendations

EPA support is recommended for the purchase by qualified reseachers of graphic systems which would permit greater development of SAR methods applied to toxicology.

MOLECULAR DESCRIPTOR 10: MOLECULAR VOLUMES AND AREAS By R. Pearlman

Evaluation of State-of-the-Art

It has been shown that solubility, partitioning and the partial dissolution whcih accompanies intermolecular interaction in solution can be discussed in terms of molecular surface areas and volumes. More specifically, theoretical considerations indicate that it is the cavity (or accessible) surface area and the van der Wall's volume which are most directly related to the free energies of the aforementioned processes (Yalkowsky 1980). Recently algorithms have been developed which enable the calculation of these areas and volumes to any desired precision with sufficient computational efficiency to permit their application to all but the largest of molecular structures.

Surface area and volume calculations are based upon well defined coordinates of the atoms comprising the molecule. Although the software exists to compute these cartesian coordinates from molecular geometric descriptors (bond lengths, bond angles, torsion angles), the preferred molecular conformation must be determined (or deduced) from experimetnal experience.

Research Needs

Many of the descriptors needed to develop QSAR for enviromental applications require the measurement or estimation of certain physical properties. There is general agreement that, for the near

107

future, most of these will have to be calculated from molecular structure. The initial success of surface area and volume based approaches to the prediction of solubilities, partition coefficients and dissolution energies suggest that it could become, at the very least, a valuable confirmation for values obtained by other procedures.

To avoid needless duplication of effort, it seems appropriate to use the newly developed software to compile a data base of surface areas and volumes for molecules of special interest to EPA. In addition, it has been suggested that these more accurately computed areas and volumes be used to construct a table of substituent group areas and volumes analagous to that published by Bondi (1964).

Recommendations

Although support for further development of software for the computer calculation of solute areas and volumes cannot be placed on a high priority basis, the EPA should encourage further studies relating these descriptors to others which have been used earlier in QSAR. It should consider making the current program available to its regional laboratories which carry out modeling and toxicity studies so that these descriptors can be conveniently included in the early development of QSARs. Only in this way can it be determined whether or not solute area and/or volume contain information missing in the descriptors in more common use.

MOLECULAR DESCRIPTOR 11: DESCRIPTOR REDUCTION By D. Weininger

Evaluation of State-of-the-Art

A large number of molecular descriptors can be generated for the purposes of developing QSARs. A common problem with the use of these descriptors is that they are often badly conditioned for use in statistical predictions, i.e., they are highly correlated with each other. Another aspect of this problem is that many descriptors which appear to be quite different actually contain redundant information. The purpose of the reduction step in QSAR work is to extract as much information into as few will conditioned descriptors as possible.

Although any and all possible descriptors may be entered into a reduction step, it is highly desirable that all descriptors are available for all compounds of interest. Primarily for this reason the only descriptors which can be considered for this approach are those calculated solely from molecular structure. It is often the case that parameters which must be measured are unavailable for many compounds (such as metabolites) and will be excluded. Topological descriptors, such as the molecular connectivity indexes, can form the basis for the pre-reduction set. Other descriptors, such as dipole moment, molecular volume, and topographic (3-D) measures of structure are also eligible for inclusion. Descriptors of bulk properties (e.g., bulk density of molar refractivity) that are calculated from some additive-constitutive relationship are suitable only if they meet the requirements of low deviation over a wide structural selection and if they can be conveniently calculated by computer.

The method of choice for descriptor reduction appears to be Principle Component Analysis (PCA). The advantages of PCA are: a large percentage (ca 99%) of the information contained in a relatively large matrix of data (ca. 100 descriptors by 1000 structures) can be reduced to a small number (5 to 10) of structural factors; the resultant factors are extremely well conditioned for linear correlative analysis; and existing PCA programs are well-proven and available. Once the PCA is done, a simple equation can be generated which allows the factor to be calculated for new compounds without redoing the PCA. The new compounds must be within the factor space of the training set, however.

Research Needs

The availability of structural factors for a large number of compounds can have two main impacts which are in the EPA's interest: The very existence of this set will promote QSAR work on the chemicals included in the set. Assuming that important industrial chemicals are included, much of the QSAR output based on the set could be directly used by EPA. Since the original descriptors are calculated solely from the molecular structure, predictions for many new chemicals could be rapidly made. QSAR's designed for screening purposes, which are based on the structural factor set, are of particular EPA interest. These screening methods could be utilized without waiting for lengthy lab tests. It should also be noted that the industrial interest would also be served. Chemicals could be 'pre-screened' for likely adverse effects before extensive efforts have gone into synthesis, production, and marketing.

Recommendations

It is recommended that EPA support an effort to compile or compute molecular descriptors which can be calculated from molecular structure, where these methods exist, and to encourage development of such descriptors The reduction of such descriptors should be done by PCA and the resultant factors should be disseminated to the QSAR resarch community.

MOLECULAR DESCRIPTOR 12: THEORETICAL CHEMICAL DESCRIPTORS bY G. Lee.

Evaluation of State-of-the-Art

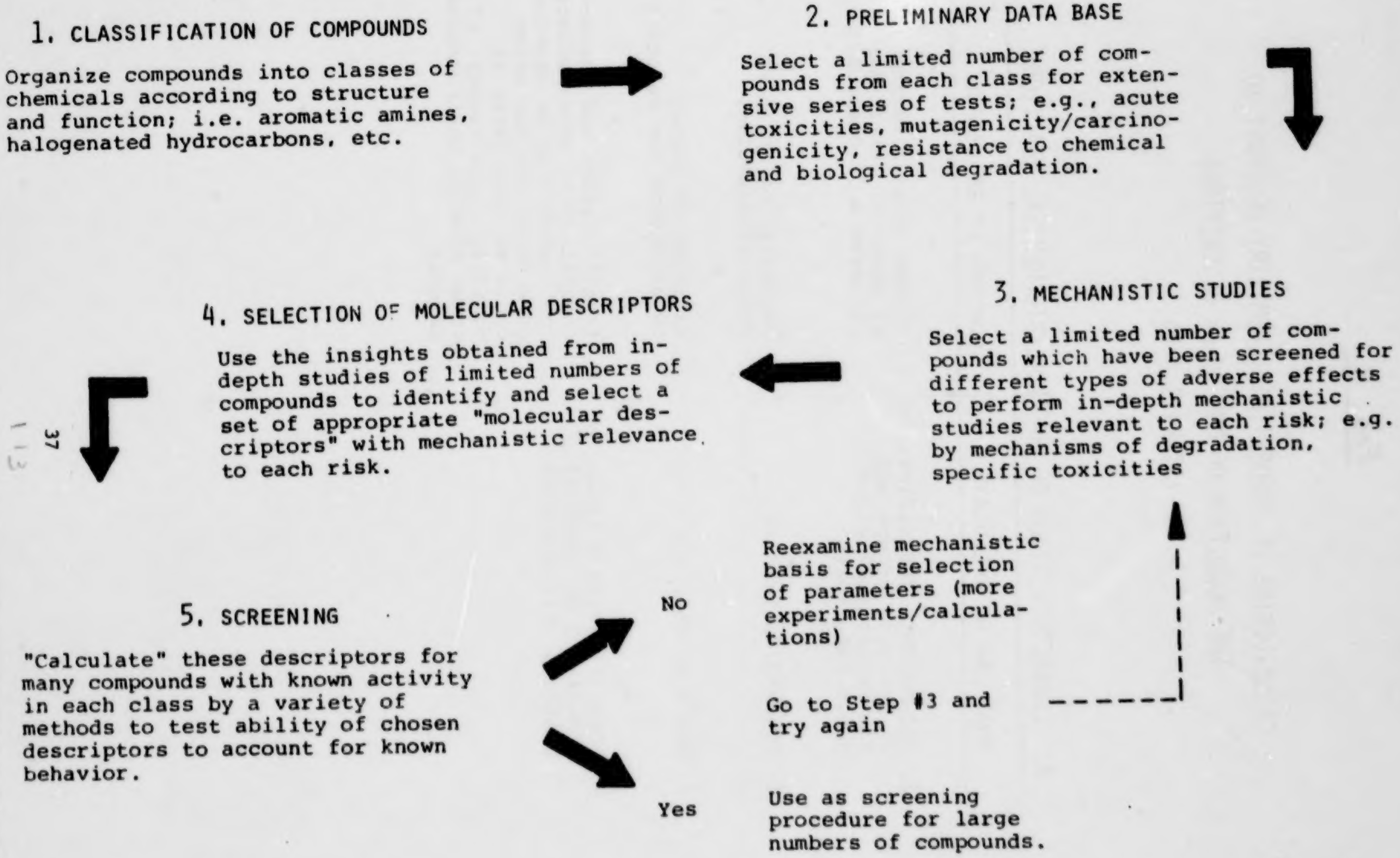
A suggested scheme for optimizing the collaboration between experimental testing efforts and 'computational' screening methods is given in Figure 1. While such studies might appear to be relatively time-consuming, they should ultimately lead to more effective and rapidly converging schemes for obtaining reliable molecular descriptors which, in turn, can given insights into mechanisms of action.

The techniques of theoretical chemistry can be particularly useful in three stages of the process outlined in Fig. 1. They can help in the elucidation of mechanisms (step 3), in the identification of mechanistically relevant (to a specific risk) molecular descriptors (step 4), and in calculation of the descriptors to be used in large scale screening (step 5).

Figures 2 and 3 summarize the information that can be obtained from theoretical chemical calcualtions and how it can be translated to identification of specific molecular properties that could be relevant to a given risk. Figure 4 summarizes how this approach can help calculate useful descriptors. Some of the methods and programs currently available are listed in Fig. 5.

Research Needs

SUGGESTED PROCEDURE FOR OPTIMAL COLLABORATION



1. CLASSIFICATION OF COMPOUNDS

Organize compounds into classes of chemicals according to structure and function; i.e. aromatic amines, halogenated hydrocarbons, etc.

2. PRELIMINARY DATA BASE

Select a limited number of compounds from each class for extensive series of tests; e.g., acute toxicities, mutagenicity/carcinogenicity, resistance to chemical and biological degradation.

3. MECHANISTIC STUDIES

Select a limited number of compounds which have been screened for different types of adverse effects to perform in-depth mechanistic studies relevant to each risk; e.g. by mechanisms of degradation, specific toxicities

4. SELECTION OF MOLECULAR DESCRIPTORS

Use the insights obtained from in-depth studies of limited numbers of compounds to identify and select a set of appropriate "molecular descriptors" with mechanistic relevance to each risk.

5. SCREENING

"Calculate" these descriptors for many compounds with known activity in each class by a variety of methods to test ability of chosen descriptors to account for known behavior.

No

Reexamine mechanistic basis for selection of parameters (more experiments/calculations)

Go to Step #3 and try again

Yes

Use as screening procedure for large numbers of compounds.

FIG. 1.

Fig. 2.

CAPABILITIES OF THEORETICAL CHEMISTRY RELEVANT TO THE SELECTION OF MOLECULAR DESCRIPTORS

(STEPS 3 & 4)

A. CHARACTERIZATION OF CHEMICAL/BIOCHEMICAL REACTIONS

WHAT CAN WE CALCULATE	HOW CAN IT BE USEFUL
1. MECHANISMS: Identification of reacting groups and the properties determing their reactivity	1. Allows identification of chemical reactivity parameters of compounds
2. ACTIVATION ENERGIES	2. Estimates of relative rates of reaction.
3. HEATS OF REACTION	3. Estimate of energy requirements for reaction
4. IDENTIFICATION & CHARACTERIZATION OF TRANSITION STATES & INTERMEDIATES	4. Identification and characterization of intermediates too unstable to be examined experimentally but which could be implicated in toxicity, persistency, or other adverse environmental effects

FIG. 3

B. CHARACTERIZATION OF NON-BONDING INTERACTIONS (COMPLEX FORMATION)

WHAT CAN WE CALCULATE	HOW CAN IT BE USEFUL
1. COMPLEX GEOMETRY	1. Identify conformational requirement for complex formation
2. COMPLEX STABILITY	2. Select electronic properties contributing to stability
3. KINETICS OF COMPLEX FORMATION	3. Estimate the importance of complex formation in the time scale of interest

Fig. 4

CAPABILITIES OF THEORETICAL CHEMISTRY RELEVANT TO CALCULATION OF MOLECULAR DESCRIPTORS: SCREENING (STEP 5)

WHAT CAN WE CALCULATE	HOW CAN IT BE USEFUL
A. ELECTRONIC PROPERTIES 1. Net Atomic Charges 2. Dipole & Higher Moments 3. Ionization Potentials 4. Electron Affinities 5. Electrostatic Potential	A. As a measure of: the extent of complex formation with polar or non-polar solvents and a variety of tissue macromolecules--DNA, proteins, and membranes
B. CONFORMATIONAL ENERGIES A set of low-energy conformers	B. As a measure of the nature and stability of complex formation with macromolecules
C. CHEMICAL REACTIVITY PROPERTIES Group or atomic electrophilicities and/or nucleophilicities	C. 1. Estimate of the relative ease and selectivity in transformation to specific intermediates 2. Estimate of the extent and specificity of covalent adduct formation with tissue macromolecules

116

FIG. 5

SPECIFIC METHODS

I. EMPIRICAL ENERGY METHODS

A. RAPID CALCULATION OF THREE-DIMENSIONAL STRUCTURES WITH OR WITHOUT GEOMETRY OPTIMIZATION

B. RAPID CHARACTERIZATION OF ENERGIES AND GEOMETRIES OF INTERMOLECULAR COMPLEXES

II. SEMI-EMPIRICAL QUANTUM CHEMICAL METHODS

A. CALCULATE ENERGY-CONFORMATION PROFILE TO OBTAIN A SET OF LOW-ENERGY CONFORMERS

B. CALCULATE ALL RELEVANT ELECTRONIC PROPERTIES

C. CALCULATE CHEMICAL/BIOCHEMICAL REACTIVITY PARAMETERS

D. MODEL CHEMICAL/BIOCHEMICAL REACTIONS

E. MODEL COMPLEX FORMATION

III. AB-INITIO QUANTUM CHEMICAL METHODS

SAME CAPABILITIES OF SEMI-EMPIRICAL METHODS, WITH GREATER ACCURACY BUT WITH MUCH GREATER TIME CONSTRAINTS

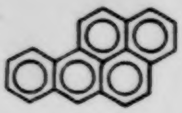
One of the least tractible problems in hazard assessment is that of a chemical which only becomes toxic after a metabolic process is initiated. If the actual toxicant appears only after several steps in the process, the molecular descriptors information obtained from the starting material is often of little or no relevance. Even though very few mechanistic studies have been completed in detail for metabolically-activated toxicants, there is already evidence that the action of cytochrome P-450 can be involved in both acute and chronic toxicity. Two examples are given in which such an in-depth study opens the way to effective QSAR using descriptors derived from theoretical chemistry. (For more complete details, see the supplemental material).

1) The current 'bay-region diol epoxide' (BRDE) hypothesis for activation of polycyclic aromatic hydrocarbons (PAH) is shown schematically in Figure 7. The first step is epoxidation of the 'distal bay region bond' by cytochrome P450 followed by hydrolysis by epoxide hydrase to a diol. This is also a substrate for P450 and the diol epoxide results. The carbocation of this diol epoxide has been implicated as the immediate precursor, forming a covalent adduct with DNA. The guanine adduct which has been identified is shown in Figure 8.

Fig. 6

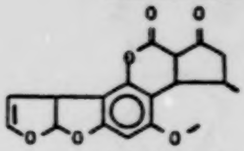
CLASSES OF CHEMICAL CARCINOGENS

POLYCYCLIC AROMATIC HYDROCARBONS



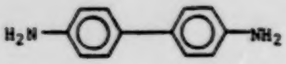
Benzo[a]pyrene
(cigarette smoke)

AFLATOXINS



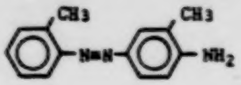
Aflatoxin B_1
(peanut butter, fungi - green plants)

AROMATIC AMINES



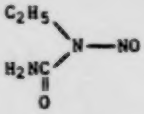
Benzidine
(laboratory reagent, dye, chemical, rubber)

AZO AND ACRIDINE DYES



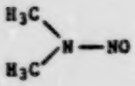
4-(*o*-tolylazo)-*o*-toluidine
(dye)

ALKYL NITROSAMIDES



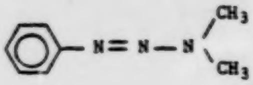
N-ethylnitrosourea
(cured meat, plastics, rubber, pesticides)

NITROSAMINES



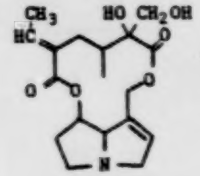
N-Nitrosodimethylamine
(cured meat, plastics, rubber, pesticides)

ARYL DIALKYL TRIAZENES



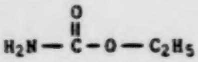
1-phenyl-3,3-dimethyltriazene
(dyes, rodent repellent, herbicide)

PYRROLIZIDINE ALKALOIDS



Retrorsine
(plants, bush teas, herbal remedies)

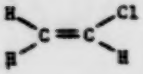
CARBAMATES



$$H_2N-\overset{\overset{O}{\|}}{C}-O-C_2H_5$$

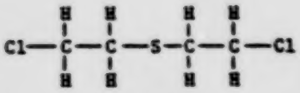
Urethan
(solvent)

HALOGENATED HYDROCARBONS



Vinyl chloride
(plastics, pesticides (DDT), refrigerant)

DIRECTLY ACTING ALKYLATING AGENTS



$$Cl-\underset{H}{\overset{H}{C}}-\underset{H}{\overset{H}{C}}-S-\underset{H}{\overset{H}{C}}-\underset{H}{\overset{H}{C}}-Cl$$

Mustard Gas
(war gas)

119

FIG. 7

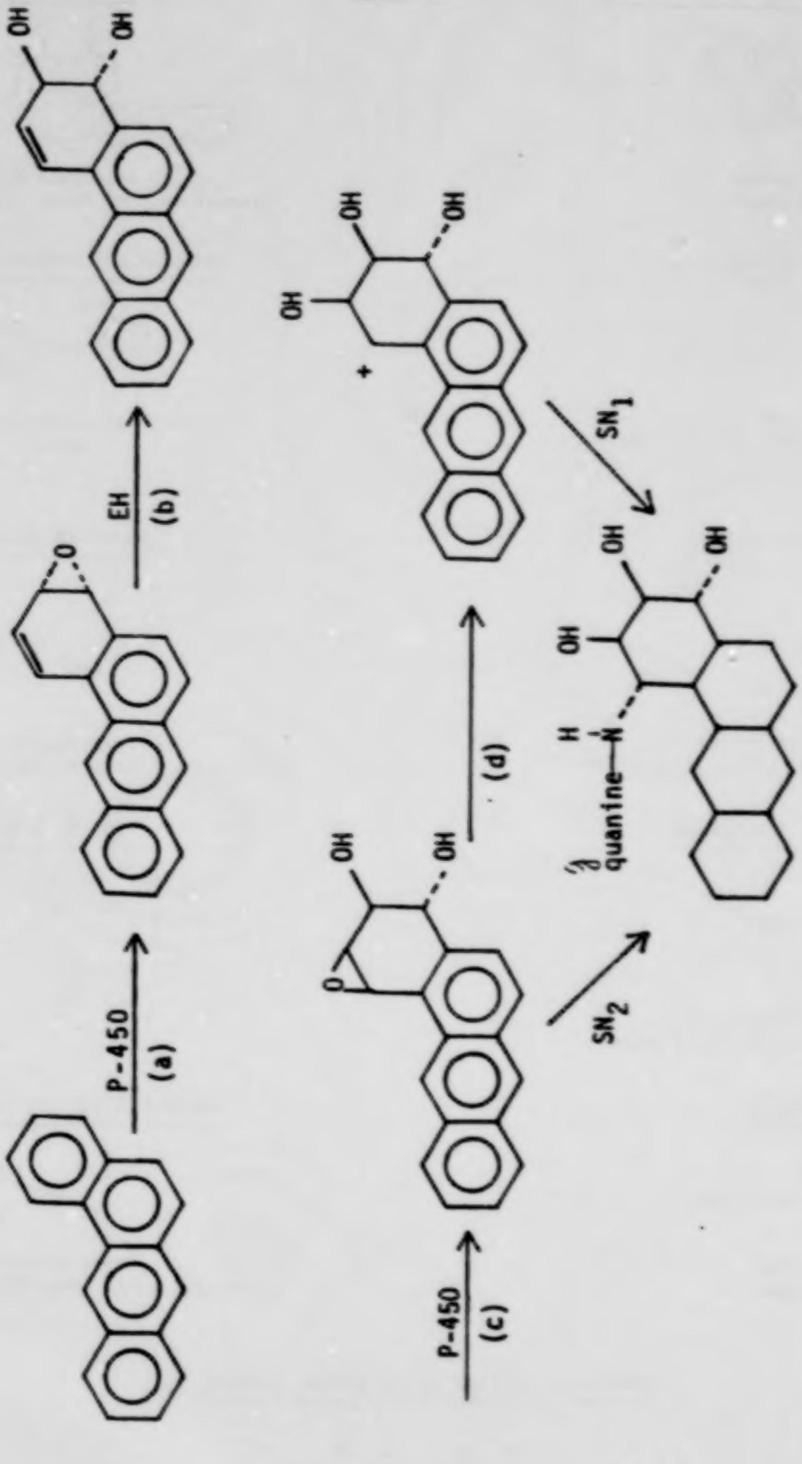
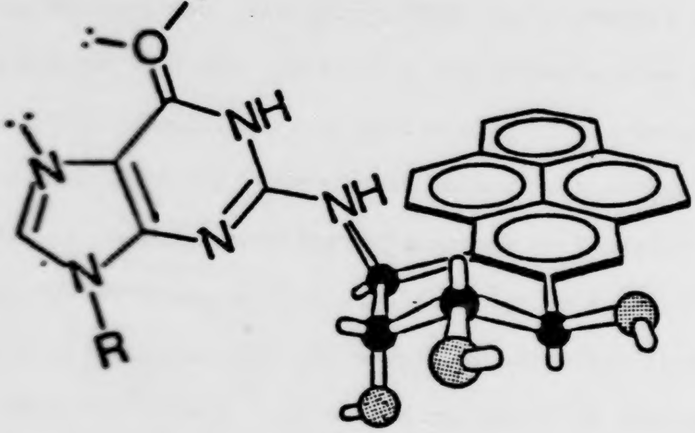

P-450
(a)
EH
(b)
P-450
(c)
(d)
SN$_1$
SN$_2$
OH
HO
O
+
H
N
guanine

FIG. 8

Structure of BP-Guanine Adduct



121

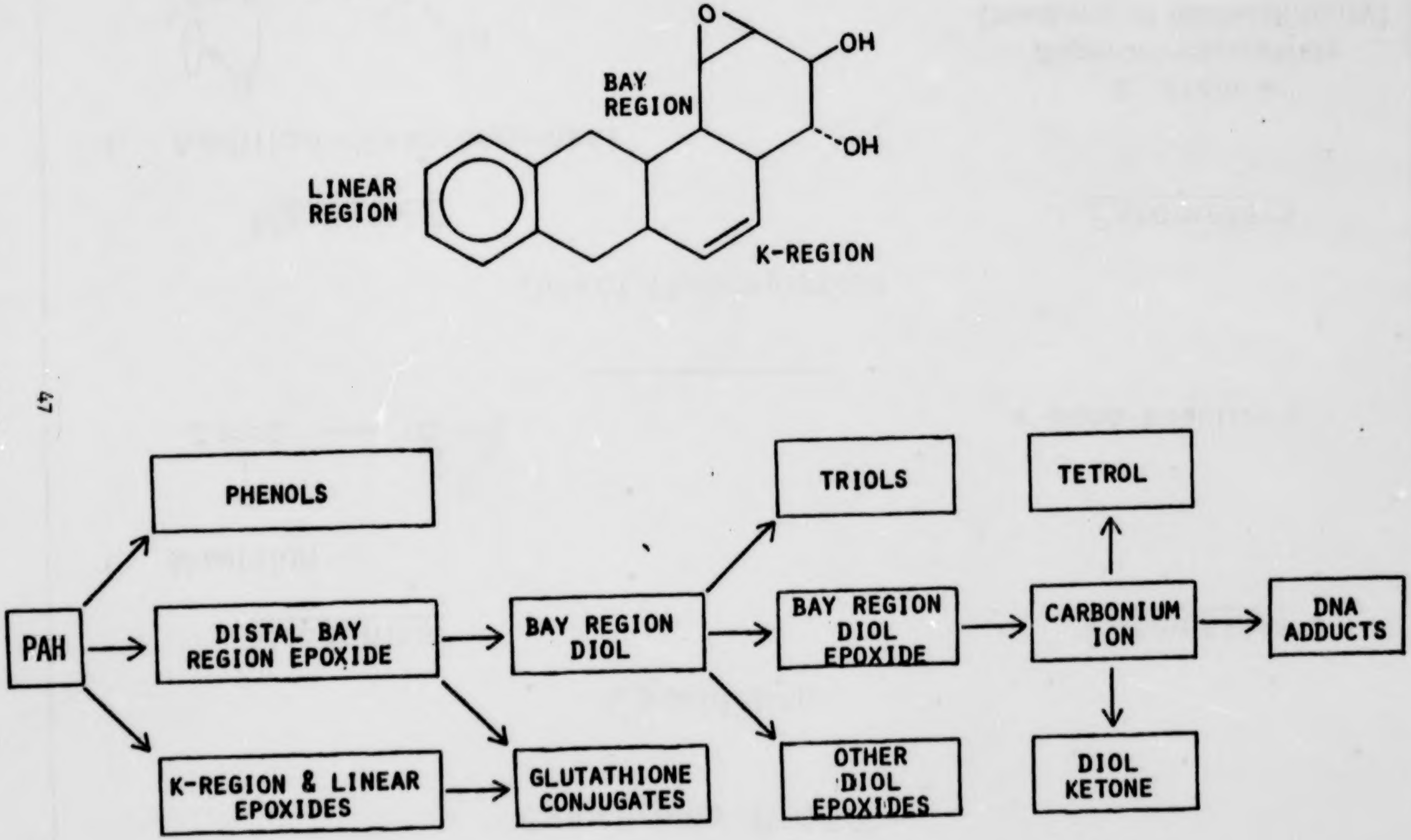
A more complete picture of the metabolic fate of the parent compound, including transformations which are now thought to lead to inactive compounds, such as phenols and other epoxides, were considered to be the ultimate carcinogens, but these are now known to be largely eliminated by other metabolic transformations. A third hypothesis had the cation radicals of the parent PAH as the active form.

In the context of the BRDE hypothesis, two types of molecular descriptors were selected and calculated, one type relevant to the extent of metabolic transformation to the ultimate carcinogen and the other relevant to the ease of formation of the BRDE carbocation. Parameters relevant to epoxidation and hydroxylation by P450 were chose as illustrated in Figure 10. Pi bond nucleophilic reactivity was calculated according to Figure 11. The molecular descriptor used to estimate ease of formation of the diol epoxide carbocation is the energy difference between it and the epoxide from which it is formed.

Table X gives the calculated energy of formation of 44 BRDE carbocations relative to the most stable carbocation. These results suggest that this single parameter could suffice as an initial screen of PAH carcinogenic activity at least to a yes or no level.

2) Some widely used haloether anaesthetics, such as methoxyflurane were found to cause significant renal and hepatic toxicity. Extensive studies showed the metabolic

Fig. 9



LINEAR REGION
BAY REGION
O
OH
OH
K-REGION
PAH
PHENOLS
DISTAL BAY REGION EPOXIDE
K-REGION & LINEAR EPOXIDES
BAY REGION DIOL
GLUTATHIONE CONJUGATES
TRIOLS
BAY REGION DIOL EPOXIDE
OTHER DIOL EPOXIDES
TETROL
CARBONIUM ION
DIOL KETONE
DNA ADDUCTS

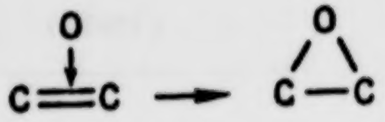
Proposed Mechanisms for Oxidation of PAH by Cytochrome P-450's

Epoxidation

Mechanism

Parameters

I. Insertion



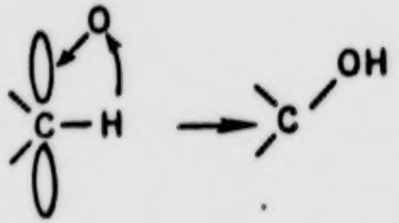
π Bond Reactivity

Direct Hydroxylation

Mechanism

Parameters

I. Addition-Rearrangement



C Atom π
Superdelocalizability
(measure of nucleophilicity)

Fig 10

Fig. 11

METHODS USED

I. CALCULATE REACTIVITIES: IEHT

π-BOND REACTIVITY

$$R_{AB} = \sum_{i\,\mathrm{FILLED}}^{\mathrm{MO}} \frac{\rho^{i}_{AB}}{\varepsilon_i}$$

ATOM SUPERDELOCALIZATION

$$S_A = \sum_{i\,\mathrm{FILLED}}^{\mathrm{MO}} \frac{q^{i}_{A}}{\varepsilon_i}$$

II. CALCULATE DIOL EPOXIDE CARBOCATION STABILITIES: INDO

$$E = (E_{\mathrm{DIOL\ EPOXIDE}} - E_{\mathrm{CARBOCATION}})$$

III. GEOMETRIES: X-RAY STRUCTURE AND TOTAL GEOMETRY OPTIMIZATION: PCILO AND MINDO/3

TABLE X

Calculated Energy of Carbocation Formation and Observed Carcinogenic Potency for 44 PAH

Compound	ΔE (kcal/mol)[a]	Carcinogenicity[b]
6,12DMBA	0	++++
Dibenzo[a,i]pyrene	1.9	++++
Dibenzo[a,h]pyrene	3.2	++++
7,12DMBA	3.8	++++
8,12DMBA	5.1	++++
12MBA	5.3	++
5,12DMBA	6.8	-
Benzo[a]pyrene	7.2	++++
5,6DMC	7.3	++ (?)
5,12DMC	7.4	-
Dibenzo[a,e]pyrene	8.1	+++
5MC	9.2	++++
6,8DMBA	9.2	++++
6-F-5MC	9.7	++++
6MBA	9.7	++
9-F-5MC	11.3	++++
7-F-5MC	11.3	++++
9,10DMBA	13.2	++
7MBA	13.3	+++
10MBA	13.7	-
Dibenz[a,h]anthracene	14.4	++
8MBA	14.4	++
9MBA	14.5	+
11MBA	15.0	-
Benz[a]anthracene	15.0	-
5MBA	15.2	-
Dibenz[a,c]anthracene	15.8	+
12-F-5MC	17.4	+
Benzo[e]pyrene	18.0	+
6MC	18.2	-
Picene	18.3	-
7,12dimethylbenzo[b]chrysene	18.3	-
2,3DMC	19.0	-
Benzo[b]chrysene	19.1	-
3MC	19.5	-
4MC	19.8	-
1MC	19.9	-
Triphenylene	20.0	-
2MC	20.0	-
Chrysene	20.6	+
Phenanthrene	22.2	-
5,10 dimethylanthracene	24.6	+
Anthracene	27.7	-
Naphthalene	35.1	-

---

[a] Energies relative to 6,12DMBA. [b] Taken from Dunning and Curtis (1960), Hecht et al. (1974), Hecht et al. (1978), Hecht et al. (1979a), Hewlett (1940), Shear (1938), Stevenson and Von Haam (1965).

50

126

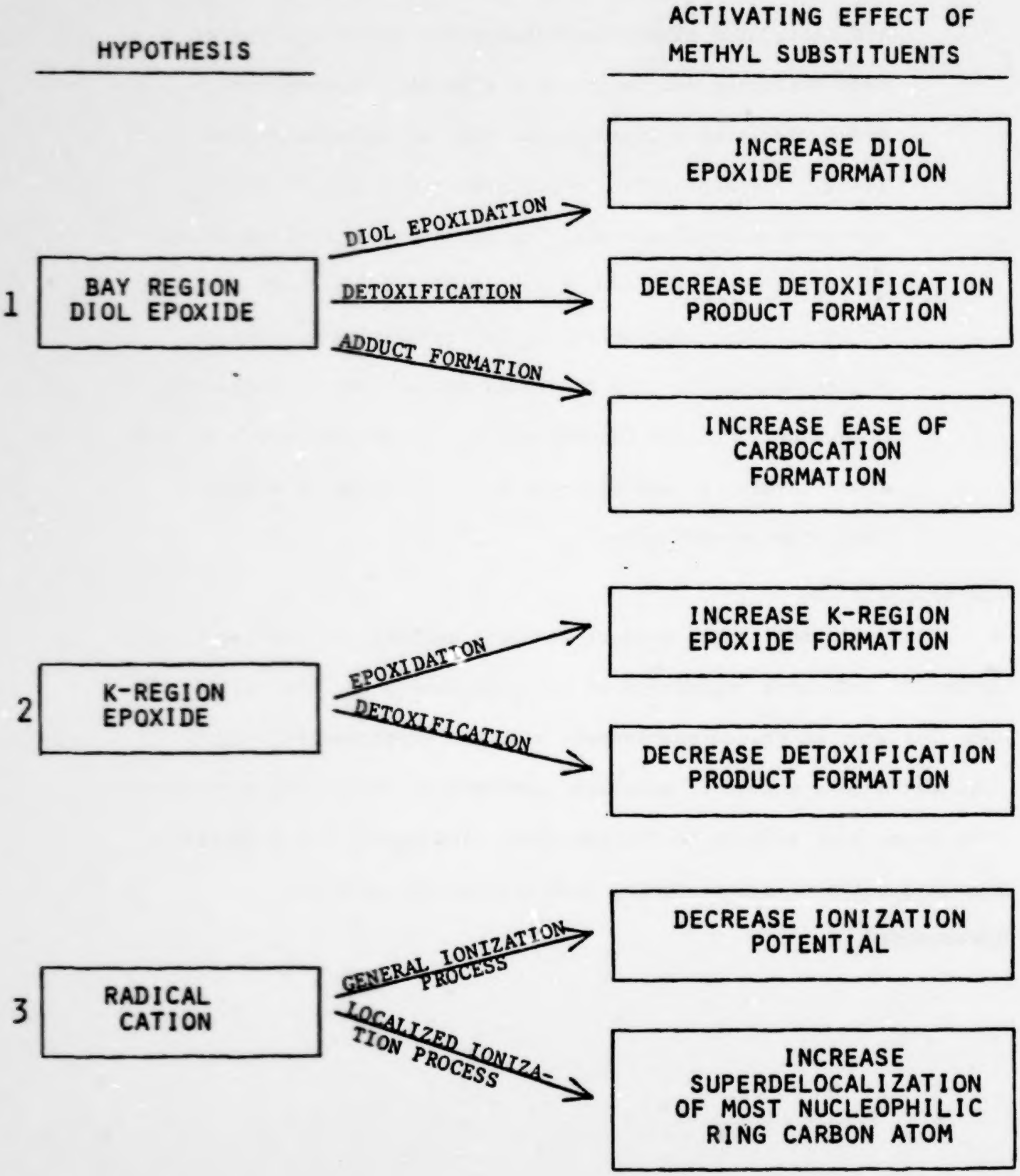
action of cytochrome P450 was responsible for the ultimate toxicant. Two plausible pathways for the metabolism of methoxyflurane and two closely related analogues involve O-demethylation and dechlorination, as shown in Figure 17-19. The appropriate descriptors for each of these pathways were calculated. The extent of metabolism of each analog as predicted from these descriptors allowed a number of significant conclusions to be drawn as summarized in Figure 22 and 23. The predicted metabolites of enflurane have been found and identified, and since they occur at much lower levels, it has replaced methoxyflurane as a general inhalation anaesthetic.

Recommendations

Obviously EPA cannot support in-depth mechanistic studies of all potential toxicants regardless of how desirable such knowledge would be. But when an ubiquitous enzyme, such as cytochrome P450 is implicated in a number of metabolic activations leading to toxicity, then support of efforts to utilize these findings in SAR compatible and supportive of other ongoing QSAR efforts is certainly recommended.

FIG. 15

THREE ALTERNATIVE HYPOTHESES FOR PAH CARCINOGENICITY


HYPOTHESIS
ACTIVATING EFFECT OF METHYL SUBSTITUENTS
1
BAY REGION DIOL EPOXIDE
DIOL EPOXIDATION
DETOXIFICATION
ADDUCT FORMATION
INCREASE DIOL EPOXIDE FORMATION
DECREASE DETOXIFICATION PRODUCT FORMATION
INCREASE EASE OF CARBOCATION FORMATION
2
K-REGION EPOXIDE
EPOXIDATION
DETOXIFICATION
INCREASE K-REGION EPOXIDE FORMATION
DECREASE DETOXIFICATION PRODUCT FORMATION
3
RADICAL CATION
GENERAL IONIZATION PROCESS
LOCALIZED IONIZATION PROCESS
DECREASE IONIZATION POTENTIAL
INCREASE SUPERDELOCALIZATION OF MOST NUCLEOPHILIC RING CARBON ATOM

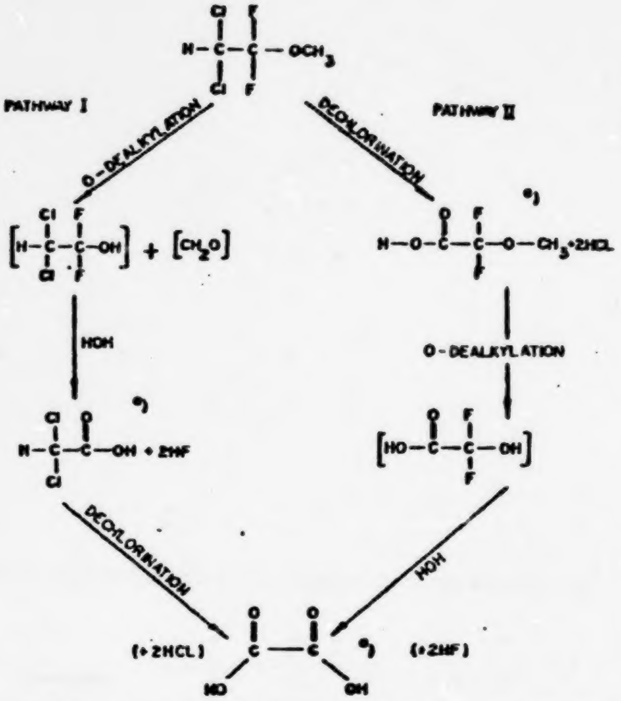

PATHWAY I
PATHWAY II
O-DEALKYLATION
DECHLORINATION
HOH
O-DEALKYLATION
DECHLORINATION
HOH
+2HCL
+2HF
(+2HCL)
(+2HF)

FIG. 17 *Proposed metabolic pathways for the anesthetic methoxyflurane*

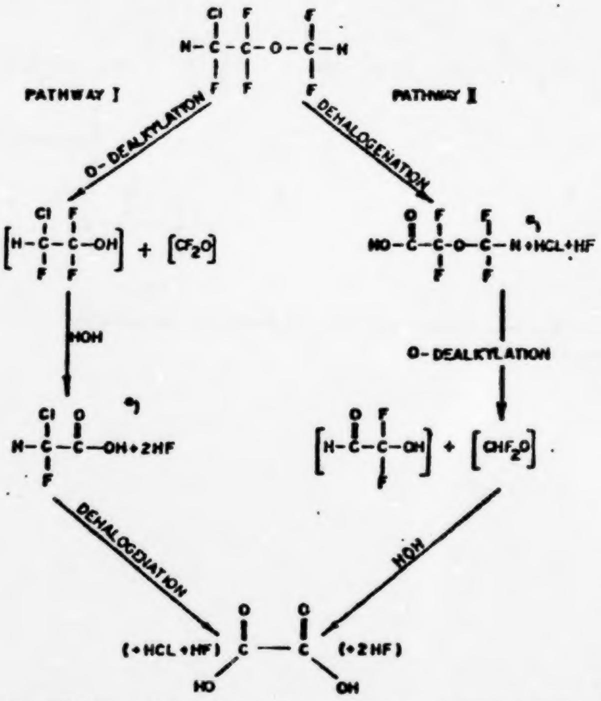

PATHWAY I
PATHWAY II
O-DEALKYLATION
DEHALOGENATION
HOH
O-DEALKYLATION
DEHALOGDIATION
HOH
+HCL+HF
+2HF
(+HCL+HF)
(+2HF)

FIG. 18 *Proposed metabolic pathways for the anesthetic enflurane*

Reproduced from best available copy.

Reproduced from best available copy.

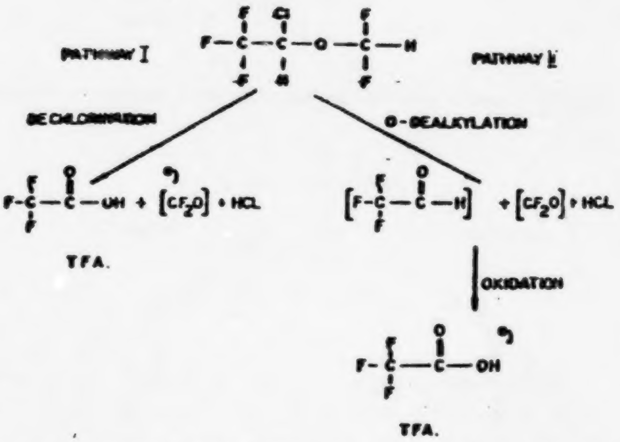

PATHWAY I
PATHWAY II
DECHLORINATION
O-DEALKYLATION
$[CF_2O]$ + HCL
$[CF_2O]$ + HCL
OXIDATION
TFA.
TFA.

FIG. 19 *Proposed metabolic pathways for the anesthetic isoflurane* TFA, trifluoroacetic acid.

Fig. 22

CONCLUSIONS

1. Overall metabolism of methoxyflurane > enflurane > isoflurane.

2. Pathway I favored for MF, Pathway II favored for EF and IF.

3. Metabolites of MF transformed by $O\text{-}CH_3$ or by dehalogenation to a greater extent than metabolites of enflurane.

4. Parent compound metabolism (Step 1) of Pathway I or II
   MF > EF > IF.

5. MF metabolized mainly by Pathway 1 Steps 1 and 2.
   EF metabolized mainly by Pathway II Step 1.
   IF metabolized by Pathway II Step 1.

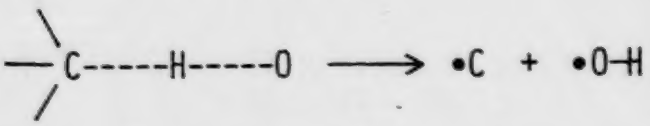
Fig. 23

CONCLUSIONS
(MECHANISTIC)

1. Dehalogenation as well as O-demethylation proceed by alkyl hydroxylation by cytochrome P450s.

2. Electrophilic O atom ("oxene") of enzyme makes a direct attack on the H atom.

    a) Rate-determining step is simultaneous C–H bond cleavage and O–H bond formation

    rather than "insertion" of O into C–H.

    b) In dechlorination, C–Cl bond cleavage might also participate in rate-determining step.

REFERENCES

Background and Summary:

1. Hammett, L.P., "Reaction Rates and Indicator Acidities", Chemical Reviews, Vol. 16, p. 67, ±935.
2. Hammett, L.P., "Physical Organic Chemistry", 2nd Ed. McGraw Hill New York, p. 347, 1970.
3. Kier, L.B., "Molecular Orbital Theory in Drug Resarch", Academic Press, New York, 1971.
4. Kier, L.B.and Hall, L.H., "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York 1976.
5. Meyer, H., Arch. Exp. Pathol. Pharmakol., Vol. 42, p. 110, 1899.
6. Overton, E., "Studien uber die Narkose", Fischer, Jena,Germany, 19Q1.
7. Randic, M., "On Characterization of Molecular Branching", J. Am. Chem. Soc., Vo.l. 97, p. 6609, 1975.
8. Taft, R.W., in "Steric Effects in Organic Chemistry", M.S. Newman, Ed., John Wiley,New York, p. 556, 1956.
9. (Yalkowsky 1980) Kier L.B., "Molecular Connectivity As A Description of Structure for SAR Analysis", MarcelDekker, New York 1980.

Partition Coefficient

1. Branson, D., 1976 (Private communiation; now published?)
2. Dunn, W., III, and Hansch, C., "Chemicobiological Interactions

and the Use of Partition Coefficients in their Correlation", Chem. -Bio. Interact., Vol. 9, 75, 1974.

3. Hansch, C. and Helmer, F., "An Extrathermodynamic Approach to the Study of the Adsorptionof Organic Molecules by Membranes", J. Polymer Sci., Vol. A-1, p. 3295, 1968.

4. Hansch, C. and Leo, A., "Substituent Constants for Correlation Analysis in Chemistry and Biology" Wiley Interscience, New York 1979.

5. Kenage, E., "Correlation of Bioconcentration Factors of Chemicals in Aquatic andTerrestrial Organisms with Their Physical and Chemical Properties", Envir. Sci. & Techn., Vol. 14, p. 553, 1980.

6. Leo,A., Hansch, C. and Elkins, D., "Partition Coefficients and Their Uses", Chemical Reviews, Vol. 71, p. 525, 1971.

7. Leo, A., "Data Base and ParameterListing" Issue #18, Pomona College Medicinal Chemistry Project, 1981.

8. Neeley, W., Branson, D. andBlau, G., "Partition Coefficients to Measure Bioconcentration Potential of Organic Chemicals in Fish", Environ. Sci. Technol., Vol. 8, p. 1113, 1974.

9. Rekker, R., "The hydrophobic Fragmental Constant",Elsevier, New York, 1977.

10. Unger,S. and Feuerman, T., "Octanol-aqueous Partition, Distribution and Ionization Coefficients ofLipophilic

134

Acids and Their Anions by Reversed-phase High-performance Liquid Chromatography", J. Chromatog., Vol. 176, p. 426, 1979.

11. Vieth, G.

12. (Yalkowsky 1980) Chou, J. and Jurs,P., "Computation of Partition Coefficients from Molecular Structures by a Fragment Addition Method", p. 163.

Solubility

1. (Yalkowsky 1980) Valvani, S. and Yalkowsky, S. "Solubility and Partitioning in Drug Design", in "Physical Chemical Properties of Drugs", Yalkowsky, Sinkula and Valvani Eds., Marcel Dekker, New York, p. 201, 1980.

Molar Refractivity

1. Hansch, 1979: see A. ref. 4.

2. Leo 1981: SEe A. ref. 7.

3. Vogel, A., J. Chem. Soc., p. 1833, 1948.

Hydrophilicity

1. Hine, J. and Mookerjee, P., "The Intrinsic HydrophilicCharacter of Organic Compounds", J. Org. Chem., Vol. 40, 1975.

B,C-Parameter

1. Cramer, R., III, "BC(DEF) Parameters. 1. The Intrinsic Dimensionality of Intermolecular Interactions in the LIquid State", J. Am. Chem. Soc., Vol. 102, p. 1837, 1980.

3.2.2.F. X-T Parameters

1. Allred, A.L. and Rochow, E.G., "A Scale of Electronegativity Based on Electrostatic Force", J. Inorg.Nucl. Chem., Vol. 5, p. 264, 1958.
2. Charton, M., Unpublished Results 1980.
3. Charton, M., "Electrical Effect Substituent Constants for Correlation Analysis", Prog. Phya. Org. Chem., Vol. 13, in press 1981.
4. Hammett, 1970: See 3.2.1. ref. 2.
5. Hansch, 1979: See A. ref. 4.
6. Taft, 1956: See 3.2.±. ref. 8.

3.2.2.H. MOlecular Connectivity

1. Hall, L.H., "Molecular Connectivity: A topologicalMethod in Medicianal Chemistry", Abstracts of Southeast-Southwest Regional ACS Meeting, New Orleans, Dec. 1980.
2. Hall, L.H., Kier, L.B. and Murray, W.J., "Molecular Connectivity II: Relation to Water Solubility and Boiling Point", J. Pharm Sci., Vol. 64, p 1974.
3. Kier, L.B., "Molecular Connectivity as a Description ofStructure for SAR Analysis", in "Physical Chemical Properties of Drugs", S. Yalkowsky, Ed., Academic Press, New York, p. 277, 1980.

4. Kier and Hall, 1976: See 3.2.1. ref. 4.

5. Murry, W.J., Hall, L.H. and Kier, L.H., "Molecular Connectivity III: Relation to Partition Coefficient", J. Pharm. Sci., Vol. 64, p. 1978, 1975.

Substructural Key Systems

1. Enslein, K.

Molecular Volumes and Areas

1. Bondi, A., J. Phys. Chem., Vol. 68, p. 441, 1964.

2. (Yalkowsky 1980) Pearlman, R., "Molecular Surface Areas and Volumes and Their Use in Structure/Activity Relationships", p. 321.

GROUP III

TABLE OF CONTENTS

Page

GROUP III STRUCTURE ACTIVITY RELATIONSHIPS(SAR) FOR ENVIRONMENTAL FATE ASSESSMENT
Background 2

Environmental Fate Process 1: Photochemical Reactions of Organic Chemicals in the Environment 9
  General Considerations 9
  SAR in Photoreaction 11
  Critique of SAR 12
  Needs 14
  References 15

Environmental Fate Process 2: Metal Complexation and Speciation 16
  Reactions of Metal Complexes and Chelates in the Environment 18
  Determination of Metal and Ligand Speciation for SAR 20
  Critique of SAR 24
  Research Needs 25
  References 29

Environmental Fate Process 3: Hydrolysis in Aquatic Systems 33
  Background 33
  Critique of SARs for Estimation of Hydrolysis Rate Constants 36
  Research and Development Needs 38
  References 44

Environmental Fate Process 4: Oxidation in Water 46
  Background 46
  SAR for Oxidation 47
  Oxidation by RO_2 47
  Oxidation by Singlet $^1O_2$ 53
  Estimation Methods 54
  Research Needs 54
  References 55

Environmental Fate Process 5: Biotransformation Processes 56
  Background 56
  Biotransformations 58
  Reaction Kinetics 60
  Biotransformation Concentration Thresholds 60

139

  - Structure Activity Relationships 62
  - Research Needs 64
  - References

Environmental Fate Process 6: Fate of Chemicals in the Atmosphere 68
  - Background 68
  - The Processes 68
  - SAR Procedures 71
  - Research Needs 73
  - References 73

Environmental Fate Process 7: Transport of Chemical Substances in the Environment 74
  - Background 74
  - Partition Data 75
  - Transport Coefficients 78
  - Research Needs 79
  - References 80

LIST OF TABLES

Environmental Fate Process 2: Metal Complexation and Speciation

1. Classification of Metals According to Relative Toxicities 16
2. Common Coordinating Groups of Monodentate and Bidentate Ligands 22

Hydrolysis in Aquatic Systems

1. LFERs FOR HYDROLYSIS REACTIONS 39

Environmental Fate Process 4: Oxidation in Water

1. OXIDANT CONCENTRATIONS IN WATER AND AIR 46
2. RATES OF OXIDATION BY RO_2o RADICAL IN SOLVENTS AT 30°C 48 48
3. RATES OF OXIDATION BY SINGLET OXYGEN IN SOLVENTS AT 25°C 48
4. SAR FOR OXIDATION OF PHENOLS BY RO_2o RADICAL IN SOLVENTS 50
5. SAR FOR OXIDATION OF AMINES BY RO_2o RADICAL IN SOLVENTS 52

Environmental Fate Process 5: Biotransformation Processes

1. Type Reactions for Transformation of Chemicals of Environmental Importance 61a
2. Type Reactions for Cleavage of Chemicals of Environmental Importance 61a

GROUP III - STRUCTURE ACTIVITY RELATIONSHIPS (SAR) FOR ENVIRONMENTAL FATE ASSESSMENT – T. Mill

Of the several hundred billion pounds of synthetic chemicals manufactured annually in this country, a significant amount enters the air, water, and soil, often at very low concentrations. Some of these chemicals persist, accumulate, and bioconcentrate, thereby affecting plant and animal life; in some cases human health is adversely affected as well.

One of the important activities of EPA is to evaluate possible environmental effects new chemicals may have if marketed and the effects existing chemicals may have because of their wide distribution, long persistence, high toxicity, or large production volumes. To facilitate this task, EPA has prepared a set of recommended test protocols for screening new chemicals for their environmental fate. The recommended protocols for fate are intended to provide data from simple laboratory kinetic or equilibrium measurements that can be used to evaluate the probable lifetime of a chemical in a specific environmental situation.

Despite the simplicity and low cost of these test methods, their widespread adoption and use by chemical manufacturers pose two major problems. First, the total cost of performing most or all of the tests may be burdensome for small manufacturers. Second, the large numbers of new and old chemicals would generate an unreasonable burden of data, evaluation of which would lead to delays and, possibly, inadequate assessments.

In many cases the testing procedure can be significantly simplified by introducing a "zero level" or prescreening step in which the rate and equilibrium constants are estimated from structure-activity relationships (SAR). In many cases prescreening can provide a reliable means of predicting the probable range of reactivity of a chemical rapidly and inexpensively.

The objectives of this study are to (1) examine the state of knowledge of SAR for environmental fate processes (2) indicate their scope and limitations and (3) recommend needed research to extend the usefulness of SAR for regulatory purposes.

BACKGROUND

Synthetic organic chemicals released into the environment are subject to movement and transformation in air, water, and soil. Metals ions, both from natural and anthropomorphic sources are mobilized or bound by natural or synthetic ligands present in these compartments. The concentrations of a specific chemical in the environment depend on how rapidly the chemical is released, where release occurs and how rapidly it moves and is transformed. Thus detailed information on sources and fate processes is an essential component of the overall hazard assessment required for regulatory purposes.

In this section we will consider SAR for the major processes and properties that control fate. Although our emphasis is on the use of SAR for estimating thermodynamic andkinetic constants, in many cases products of transformations can be predicted quite accurately (as for example in hydrolysis).

At present we believe that the following processes can be important in soil, water, and air:

SOIL/SEDIMENT	WATER	AIR
biotransformation	metal complexation	oxidation
hydrolysis	hydrolysis	photolysis
surface photolysis	oxidation	deposition
volatilization	photolysis	
sorption	sorption to sediments	
reduction	biotransformation	
biouptake	biouptake	

Movement among compartments is controlled both by equilibrium and kinetic processes, but estimates based upon thermodynamic properties can give useful first approximation to distribution that will assist regulators and manufacturers in focusing attention on those compartments where a significant fraction of the chemical would accumulate. This distribution estimate provides only a boundary condition and in many instances more detailed estimates which include kinetic parameters will indicate that nonequilibrium conditions actually apply as for example in movement from soil or water to air.

The underlying assumption in use of kinetic and equilibrium process data is that the net movement and loss of a chemical is the simple sum of all such processes, each of which is expressed, for simplification, as a first-order rate relation

$$d\,[C]\,/dt - -k\,[C] \qquad (1)$$

where $[C]$ is the concentration of chemical and k is a pseudo-first order constant which combines both environmental and structure parameters. That is

$$k = k'[P]$$

where $[P]$ is the environmental property such as photon flux, acidity, or basicity, or sediment carbon content. If input is continuous, the net concentration of chemicals at any time will be in a steady state resulting from the balance of input and loss rates. Therefore decisions about what constitutes a significantly fast or slow rate of loss must be tempered by consideration of the rate of input; for example, a chemical which hydrolyzes rapidly (half-life of a few days) may still be present in an aquatic system at significant concentrations if the input rate is high.

SAR can be a valuable means of eliminating certain processes from a detailed fate assessment if the value of the estimated rate or equilibrium constant is so small that the process cannot compete with other faster processes. Under these circumstances. The precise value of the rate constant is not important. Processes which have half-lives of many years or many hundred of years generally will not control fate, and more reliable estimates of the rate constants are unnecessary. Thus use of both qualitative and quantitative SAR can greatly simplify fate testing procedures.

The following sections provide more detailed discussions of the important fate processes, the state of knowledge of SAR for estimating the process constants, gaps in our current understanding, and recommended research needs to improve the reliability and scope of SAR.

We have not addressed the processes of uptake by plants of chemicals from soil or water, dry deposition of chemicals from the atmosphere, or reduction in soils/sediments. All may prove to be important mechanisms for concentration or movement but none appears to be well enough understood at this time to provide any basis for SAR.

Structure activity relationships (SAR) can greatly simplify the task of estimating rate and equilibrium constants needed for modeling transport and transformation processes in water, air, and soil. SAR can also assist in selecting best values from experimental measurements. Because the predictive power of SAR varies widely (factors of two to orders of magnitude), clear indicators of their accuracy and limitations are essential to avoid misapplication and unreasonable extrapolations and expectations.

The best developed SAR are for transport and chemical transformation constants in water and air. Thus solubility, sorption, and diffusivity can be estimated quite precisely from structural parameters. Hydrolysis in water and oxidation in water and air are wellunderstood processes, but SAR for these processes are limited by inadequate data bases. Metal complexation by natural and synthetic ligands are controlled both by thermodynamic and kinetic parameters; some SAR are available for metal ions and synthetic ligands; however, sorption and ion exchange with clay minerals may confound these estimates in unknown ways.

Photolysis in water and air often is complex, and their rate constants can be estimated only as upper limits. For biotransformations only broad generalizations and upper limits for rates can be given at this time.

Except for sorption, none of the processes believed to be important in soils (such as hydrolysis, biotransformation and uptake) are well understood and only very limited SAR are available at this time.

The following section summarizes major research needs identified by Group III as essential for additional development of useful and reliable SAR for environmental processes; needs are ranked somewhat by overall assessment of importance and likely early progress.

Biotransformation in Water or Soil - Develop a better understanding of the molecular basis for acclimation and transformation, including hydrolysis, oxidation, and reduction in order to more accurately measure rate constants for biotransformation at low concentrations of chemicals, and from this information develop SAR for the reactions.

Soil Processes - We need to understand how soil humus and clay minerals interact with chemicals, and can either catalyze or inhibit hydrolysis, oxidation reduction, photolysis, and volatilization in order to provide the mechanistic basis needed for development of SAR.

Hydrolysis in Water - Extend the kinetic data base for hydrolysis in pure water for more classes of structures over a useful pH range (3-11); explore structure-mechanism relations in water and application of existing SAR developed in mixed solvents.

Photolysis in Water and Air - Extend the data base for quantum yields for photolysis of structurally similar compounds; explore how UV spectra and quantum yields measured innonpolar solvents might be used in water or air.

Oxidation in Air - Extend addivitity SAR for oxidation by HO radical to more substituted alkanes, olefins and aromatics; explore development of SAR for oxidation by ozone.

Oxidation in Water - Develop data base for oxidation of reactive structures by RO_2 radical and singlet oxygen in water; examine the applicability of existing SAR based on reactions in organic solvents.

Volatility - Develop SAR for vapor pressures of high molecular weight chemicals based on critical volumes and boiling points.

ENVIRONMENTAL FATE PROCESS 1: PHOTOCHEMICAL REACTIONS OF ORGANIC CHEMICALS IN THE ENVIRONMENT by R. Srinivasan

General Considerations

The only source of light energy for photochemical activity in the environment is the sun. In the troposphere and in water, useful solar radiation of photochemical interest extends from ~290 nm to ~730 nm. The distribution in the intensity is not uniform over this region and drops sharply at wavelengths shorter than 400 nm which is the region of strong photochemical activity. Photon density at the surface of the earth is readily available in the literature (Mabey et. al., 1979; Zepp and Cline, 1977).

Photochemical reactions can be brought about in one or two ways:

(1) A compound can absorb the light and undergo a direct chemical transformation unimolecularly, or bimolecularly with the medium (e.g., water or oxygen).

(2) A compound can absorb the light and initiate the reaction of another molecule by sensitization while the abosrber itself is not destroyed. (e.g., photooxidation in which atmospheric oxygen is activated by a dye.)

More complex reaction sequences are also known. For example, photolysis can lead to free radicals which can set up a thermal chain reaction involving the substance itself or the medium. The best example of such reactions is in the generation of photochemical smog. This subject will be dealt with in the section on air oxidation.

Direct photolysis rates in sunlight at low concentration of a chemical can be formulated as a function of the solar irradiance (L_λ), absorptivity (ϵ_λ), quantum yield (ϕ), and the concentration of the chemical by the equation 3.1:

$$\frac{-d[C]}{dt} = 2.3\,\Sigma(L_\lambda \epsilon_\lambda)\,\phi\,[C] \qquad 3.1$$

This form of the equation is applicable only when $\epsilon[C] \leq 0.02$. In the case of a photosensitized reaction, the form of this equation would not change because the sensitizer would absorb the incident light at a constant rate and the destruction of the chemical of interest would depend only on its concentration. In this instance, the equation can be written as in 3.2

$$\frac{-d\,[C]}{dt} = L_\lambda x\,\phi[C] \qquad 3.2$$

where 'x' represents a constant fraction of the incident light that is absorbed by the sensitizer at a given wavelength, and is the quantum yield for the sensitized reaction.

Sar in Photoreaction

Only light that is absorbed by a molecule is capable of bringing about a photochemical reaction. The ultraviolet and visible absorption spectrum of a compound in the region of the solar radiation is an index of its potential for direct photoreactivity. The chromophores in a molecule which determine its absorption spectrum are well recognized and tabulated extensively (Calvert and Pitts, 1967; Landolt-Bornstein 1951; Perkampus, 1971).

Rate data on the photochemical decomposition of a variety of compounds with simple chromophoric groups are available in the literature and compiled in one publication (Calvert and Pitts, 1967). In most instances, quantum yields at one wave length are given. From this information by the application of equation 3.1, it is possible to give an estimate of the halflife of the chemical in air or water under a standard solar radiation intensity.

Compounds with many chromophoric groups in conjugation have extended absorption spectra. Such compounds may photolyze at the most labile center, which can be judged by an inspection of the structure. There is an implicit assumption that the compound may undergo principally a cleavage reaction. The quantum yield for decomposition is more difficult to estimate under these circumstances because extended chromophoric systems are usually efficient at degrading and dissipating the photon energy. Thus the quantum yield

152

for the destruction of the chemical may be much lower than for one with a nonconjugated system.

In making quantitative estimates of the fate of a chemical which degrades by a photosensitized process, the application of equation 3.2 requires a knowledge of the structure, concentration, and absorption spectrum of any sensitizing species which may be present in the environment. In water, the active material is taken to be humic acid at a concentration of 1-6 mg/l. If the quantum yield for the sensitized process is available, an estimate based on equation 3.2 of the half-life of the chemical in question becomes possible. For a given sensitizer (e.g., humic acid), the quantum yield for the sensitized decomposition of a series of homologous compounds can be taken to be of the same order of magnitude.

Critique of SAR

To relate structure to activity is enormously more difficult in photochemical reactions than in thermal reactions. The explanation is simply that the reactant in the photochemical process is the elecronically excited state of the molecule which can be looked upon as a new, transient chemical compound whereas the structure one examines is that of the ground state. Attempts to extend quantitatively the relationship between sructure and photochemical reactivity in a series of homologous compounds under controlled conditions (wavelength of light, temperature, solvent) have invariably given rise to *very* limited correlations (Calvert and Pitts, 1967). For example, in a series of aliphatic ketones of the general structure $RCOCH_3$ where R is an alkyl group, although the electronic transition is the same in the near-ultraviolet region, photodecomposition proceeds by α-cleavage in the cases in which R is methyl or ethyl and predominantly by a type II process when R is n-propyl or n-butyl. This can be attributed, as already pointed out, to the differences in the distributions of the electrons in the excited states of the homologous compounds, but it certainly does not encourage SAR. The optimum strategy would be to keep the extrapolation from the known to the unknown to an absolute minimum.

Practical problems in the use of SAR also exist. Chemicals in the environment are found as dilute solutions in water or air. Data in the literature on both absorption spectra and photochemical processes rarely extend to water or the gas phase. An extrapolation of the data from one solvent medium (e.g., a hydrocarbon) to another (water) may be unwarranted. Indeed, it is known that for many chromopores, a change from a nonpolar to a polar solvent causes shifts in the absorption spectra as well as changes in the photochemistry. Data on the photochemistry of organic compounds in aqueous media and in air are quite limited although these are the most relevant and essential for the present purpose.

Organic photochemistry has grown extensively over the past 25 years, but there are still large gaps in our knowledge which make the prediction of the reactivity and the nature of the product for many compounds essentially a guess.

Needs

Photochemistry of organic systems in water or air is largely unexplored. The principal experimental problem in such studies is the diluteness of the sample which is caused by the volatility of the compound in air or its solubility in water. However, these difficulties are not insurmountable, and such studies should be urgently pursued in order to minimize the extrapolation that is involved in estimating the reactivity of a new compound.

REFERENCES

Calvert, J.G., and J. N. Pitts, Jr., 1967. Photochemistry. John Wiley, New York.

Landolt-Bornstein. 1951. Zahlenwerte andFunktionen. Springer-Verlag, Berlin.

Mabey, W.R., T. Mill, and D.G. Hendry. 1979. Test Protocols for Environemntal Processes: Photolysis in Water. EPA Report. EPA Contract 68-03-2227.

Perkampus, H.H., et. al., 1971. UV Atlas of OrganicCompounds. Vol. I-V. Plenum Press, New York.

Zepp, R.G., and D.M. Cline. 1977. Rates of Direct Photolysis in Aquatic Environments. Environ. Sci. and Technol. 11: 359-366.

ENVIRONMENTAL FATE PROCESS 2: METAL COMPLEXATION AND SPECIATION

A.E. Martell

Introduction

The metal ions and metal compounds found in the environment may be classified as relatively nontoxic; toxic and accessible; and nonaccessible, but toxic, as indicated in Table 1 (Wood, 1974; Martell, 1975).

TABLE 1. CLASSIFICATION OF METALS ACCORDING TO RELATIVE TOXICITIES

Non-Toxic	Toxic and Accessible	Non Accessible, but Toxic
Li, Na, K, Rb	Be, Zn, Cd	Ti, Zr, Hf
Mg, Ca, Sr	Co, Ni, Cu	Zr, W, Mo
Al, Fe, Si	As, Sb, Bi	Nb, Ta, Re
	Se, Te, Sn, Pb	Ga, La, Lanthanides
		Os, Rh, Ir
		Ru, Ba, Actinides

Reproduced from
best available copy.

A general classification of this nature must be interpreted with care, since in most cases toxicity depends on both the level (concentration) and speciation of the metal ion of concern. For example, small amounts of Zn, Co, Mn, Cu and Mo are essential to life because they are required for enzyme function. Also, other "toxic" or carcinogenic metals such as Ni, Cr, V, Pb, and Sn are now considered possible candidates as beneficial or even essential trace elements for enzyme systems (Schwarz, 1974; Nielsen, 1974, 1980). Similarly, metal ions listed as nontoxic such as Al and Fe are known to be highly toxic when they accumulate in excessive amounts in physiological systems (Eichhorn et. al., 1980; Anderson, 1980). Also, certain "noble" metals which are generally in the environment as inaccessible minerals, may be converted by chemical or microbiological processes to highly toxic and available forms. Such processes have been well documented for mercury (Wood and Brown, 1972; DeSimone, et. al., 1973).

The mobilities, properties, and toxicities of organic ligands (organic compounds that coordinate metal ions) are strongly influenced by metal ion complexation. For example, EDTA, which has a high affinity for the calcium ion, is for that reason highly toxic when introduced into physiological systems as the sodium salt (Saffer and Toribora, 1961). When injected as the calcium chelate, however, it is harmless and may be used in that form for chelation therapy for the removal of lead and for other heavy metal intoxications (Rubin, 1953). Such considerations are not of critical importance for EDTA in the environment because it is not absorbed through the gut. On the other hand, there has been some concern about NTA, which is partially absorbed from the intestinal tract. The fact that the toxicity of NTA in the environment is lowered by metal complexation has been pointed out (Rubin and Martell, 1980), and NTA has been shown to have potential medical applications through oral administration (Rubin, et al., 1980).

From the above, it is obvious that the toxicities, as well as the mobilities and other properties of organic ligands and metal ions are profoundly influenced by complexation. Therefore one of the requirements for applying SAR to environmental systems containing metals is to determine the speciation of the metal compounds present. Thus, methods should be worked out that will make it possible to estimate the interactions of 20-30 metals (naturally occurring, and those that may be introduced through environmental pollution) and the thousands of potential organic pollutants that contain functional groups that may chelate or complex the metals of interest.

Reactions of Metal Complexes and Chelates in the Environment

Metal ions in aqueous solution are generally highly solvated and may be considered complexes of water, so that metal complex or chelate formation may be considered the replacement of one ligand (water) by another. Such reactions may be very rapid, or moderately rapid, as in the case for most accessible (see Table 1) metal ions of interest, but may be extremely slow for so-called "inert" metal ions such as Cr(III) and Co(III). Thus complex formation or ligand exchange of chromium complexes may take days or months to reach equilibrium at ambient environmental temperatures. For different reasons, very stable metal chelates may exchange very slowly, as for example, the conversion of Cu(II)-EDTA to a Cu(II) porphyrin at environmental temperature and pH. It may be seen therefore, that any attempt to predict metal speciation must take kinetic factors into account.

Metal ions may also catalyze reactions of the organic ligands to which they are coordinated. Thus, metal-catalyzed hydrolysis of esters, amides, and of other organic functional groups has been widely studied (Martell, 1968). Oxidation of organic ligands by various oxidizing agentse are also well-known (Martell, 1968; Lappin, 1977) but informtion in this area is much less complete. Of special importance for the environment is the oxidation of ligands by molecular oxygen as the oxidant. Such reactions are frequently photocatalytic, and many ligands are rapidly oxidized by metal ions in the presence of sunlight. The current incomplete state of the art of metal-catalyzed oxidation reactions, and particularly of photochemical oxidation reactions, constitutes a serious deficiency in the data needed for the estimation of metal and organic ligand speciation as a function of time under oxidative conditions.

Another area for which scientific data are urgently needed involves adsorption and ion exchange of metal ions and complexes in solution by mineral deposits present in the environment. While distribution of metal ions between water and certain clays and mineral ion exchangers has been measured, little is known about the adsorption and exchange of metal chelates and of organic compounds containing complexing functional groups. Such interactions can profoundly influence the levels of organic substances as well as of metals in the environment. Adsorption processes have been observed to reverse the predictions of metal solubilization based on metal chelate stabilities alone, as in the case of iron-NTA systems (Martell, 1975).

Determination of Metal and Ligand Speciation for SAR

The compilations now available on metal ion-ligand affinities (Sillen and Martell, 1964, 1971; Smith and Martell, 1975, 1977; Martell and Smith, 1974, 1976) make possible the calculation of concentrations of metal complexes and chelates for aqueous systems containing a wide vareity of metals ions and hundreds of organic ligands with a considerable degree of confidence.

Similar calculations may be carried out for the thousands of organic environmental pollutants with coordinating functional groups by estimating their metal ion affinities through comparisons of their chemical structures with those of ligands that have been measured. While such estimates are not as accurate as the concentrations that would be determined by direct experimental measurement of stabilities, predictability in most cases would be within a factor of 2 or 3. It would be impossible to measure the affinities of the thousands of ligands that may occur as environmental pollutants for the 20-30 metal ions that occur naturally or as pollutants in environmental systems. A further complication that points up the impossibility of direct measurement of all possible metal species is the formation of mixed complexes involving the combination of a metal ion of coordination numbers of 4, 6, or more with two or more different ligands having a small number (1-3) of coordinating functional groups. Once the equilibrium constants for simple metal-ligand interactions have been measured, or estimated from known model compounds of similar structures, the stabilities of the mixed complexes may be calculated on the basis of statistical considerations. Fortunately the probability of mixed ligand complex formation is greatly reduced in dilute solution, thus simplifying the speciation and the calculation for most environmental systems.

Estimates of metal affinities of ligands that have not been previously measured may be carried out in a manner analogous to the following examples. Standard compounds may be found in the stability tables cited above for organic compounds containing monofunctional (i.e., monodentate) coordinating groups of the type listed in Table 2, as well as of bidentate combinations of two of these monodentate donor groups.

TABLE 2. COMMON COORDINATING GROUPS OF MONODENTATE AND BIDENTATE LIGANDS

Common Monodentate Donor Groups	Bidentate Combinations	
1. Amino $R\text{-}NH_2$	Amino Acid	1+2
2. Carboxylate $R\text{-}COO^-$	Hydroxy Acid	3+2
3. Aliphatic hydroxyl R-OH	Phthalic Acid	2+2
4. Phenolate $R\text{-}O^-$	Organic diamine	1+1
5. Mercaptide $R\text{-}S^-$	Catecholate	4+4
	Salicylate	2+4

Each of the monodentate donor groups and bidentate combinations listed are represented by dozens of standard ligands in the stability constant tables. Affinities of a new ligand for various metal ions may be estimated from the values given for the ligands that resemble most closely the ligand of interest. In many cases there are sufficient examples to provide trends in stability constants based on structural variations in the ligands.

In addition to estimating the metal-ion affinity constants, it is also necessary to estimate the basicities (i.e., pK's) of the ligand functional groups, so that hydrogen ion competition at environment pH values can be calculated. For metal ions that readily hydrolyze, such as Fe(III), Al(III), Cr(III), La(III), etc., it is necessary to take into account metal ion hydrolysis, as determined by K_{sp} values, and in some cases to estimate hydrolysis constants of the metal chelate or complex, whereby the hydroxide ion and the ligand are simultaneously coordinated to a metal ion.

In the estimation of formation, hydrolysis, and protonation constants for new ligands it is necessary to utilize the guidelines that relate the nature of the functional groups and their steric arrangements in the molecule that contribute to metal ion and hydrogen ion affinity. These relationships have been studied by coordination chemists and have been described in recent reviews (Ahrland, 1968; Martell, 1960, 1978, 1981a, 1981b).

Critique of SAR

Because of the ability of metal ion-organic ligand systems to undergo exchange with competing ions and ligands in the environment, metal complexes may readily undergo changes of speciation when their concentrations change in the environment as the result of dilution, metal-catalytzed ligand reactions, or exchange with mineral deposits. Thus an organic chelating agent released into environmental waters may undergo extensive changes of toxicity (usually reduction of toxicity) as its speciation changes from that of a free ligand to the chelates of the naturally occurring metal ions. It has been pointed out (Rubin and Martell, 1980) that the toxicities of chelating agents and their metal chelates may differ by several orders of magnitdude, and that the chelates of metal ions that are naturally present in biological systems generally have the lower toxicities.

Similar changes in the toxicities of metals with speciation, which vary widely with human dosage or exposure level, have also been observed. It is well known that many forms of Cr(III) and Ni(II), including many simple complexes, are highly toxic and carcinogenic in man when exposure occurs at moderate to high concentrations. Similarly, copper(II) compounds have been found to be co-carcinogens (Martell, 1981b and references therein) at moderate concentration levels. On the other hand, all three metal ions have been found to be essential to life at trace levels because they are required for the activation of enzymes. Thus administration of these metal ions at trace concentrations is not harmful since they would be complexed by naturally occurring ligands to form metal transport complexes and metal-activated enzymes. When these metals are supplied in excessive amounts, however, the natural metal receptors are overwhelmed, and the excess metals may then have toxic effects on other biomolecules that have coordinating donor groups.

For the above it is obvious that metal complex speciation may change drastically with concentration in both environmental waters and in biological fluids, because of competition with low but often significant levels of naturally occurring metal ions and ligands. Thus it is generally not possible to extrapolate toxicity data measured at high concentrations to low environmental or biological levels. This behavior contrasts to that of organic molecules that do not change composition on dilution, and for which the extrapolation of toxicity data to low concentrations is sometimes considered valid (See Group IV).

Research Needs

1. Speciation

In order to estimate metal speciation in the presence of environmental organic pollutants, the following research plan is recommended:

a) By careful inspection of the 50,000 chemical substances that are potential environmental pollutants, select those that have complexing functional groups and classify them according to type, number, and steric arrangement of their coordinating groups.

b) By comparison with analogous complexing agents for which proton and metal ion affinities have been determined, estimate the pK's and metal ion stability constants of the compounds under consideration. In this manner, set up a compilation of equilibrium constants for all complexing pollutants of interest.

c) Test the accuracy of the speciation calculations based on the complilation of estimated equilibrium constants by (1) direct experimental determination of metal chelate stabilities of various types of representative ligands, and (2) by the use of especially designed "pilot" ligands that will allow direct measurement of metal speciation and concentration of complex mixtures.

d) Apply the data compilation to the determination of metal complex concentrations in environmental waters containing mixtures of several ligands and metal ions. Predict toxicities of the mixtures on the basis of the speciation and toxicities of the compounts, and verify by direct measurement of toxicities of the mixtures with experiemental animals.

2) Catalysis

On the basis of known examples of metal-catalyzed hydrolysis, oxidation, and photooxidation of organic pollutants, predict the conversion of organic substances from noncomplexing to complexing ligands. With the chelating agents (e.g., carboxylates, phenols, catechols, etc.) thus generated, set up a new compilation of organic ligands and their metal chelates, using the procedures outlined in (1) above. With known kinetic parameters of model compounds, estimate the variations with time of metal speciation and metal-ligand concentrations for complex systems containing several metal ions and organic ligands. The estimated levels and speciation should be calibrated by using standard ligands and observing rates of change of their structures, speciation and concentrations.

3. Exchange with Minerals and Soils

Set up a number of standard soils, clays, and minerals of the type frequently encountered in the environment, and determine the redistribution of the metal ions, ligands, and complexes between the aqueous and solid phases. Classification of distribution ratios according to structures and types of metal complexes would then be used to estimate distributions for other metal ligand complexes and to estimate the changes in speciation that would occur in complex systems. Calculations of the redistributions should then be applied to modify the speciation estimates of the type described in (1) above. The resulting distributions should then be checked for real systems in which the speciation of individual metal complexes can be measured experimentally.

References

Ahrland, S., Thermodynamics of Complex Formation between Hard and Soft Acceptors and Donors, Structure and Bonding, 5, 118-149 (1968).

Anderson, W.F., Iron Chelation in the Treatment of Cooley's Anemia, in: "Inorganic Chemistry in Biology and Medicine" ACS Symposium Series 140, Ed. Martell, A.E., American Chemical Society, Washington, D.C., 1980, pp, 252-261.

Eichhorn, G.L.; Butzow, J.J.; Clark, P.; von Hahn, H.P.; Rao, G.; Heim, J.M.; Tarien, E.; Crapper, D.R.; and Karlik, S.J., Metal Ion-Nucleic Acid Interactions: Aging and Alzheimer's Disease, in: "Inorganic Chemistry in Biology and Medicine: ACS Symposium Series 140, Ed. Martell, A.E., American Chemical Society, Washington, D.C., 1980, pp. 75-90.

Lappin A.G., Metal-Ligand Redox Reactions, in: Special Reports: Inorganic Reaction Mechanisms, Chemical Society, London, 1977, pp. 42-106.

Martell, A.E., The Relationship of Chemcial Structure to Metal-Binding Action, in: "Metal Binding in Medicine", Ed. Sevin, M., Lippincott Co., Philadelphia, 1960, pp. 1-18.

Martell, A.E., Catalytic Effects of Metal Chelate Compounds, Pure & Appld. Chem., 17, 129-178 (1968).

Martell, A.E., The Influence of Natural and Synthetic Ligands on the Transport and Function of Metal Ions in the Environment, Pure & Appld, Chem., 44, 81-113 (1975).

Martell, A.E., Chelating Agents for Metal Buffering and Analysis in Solution, Pure & Appld. Chem., 50, 813-829 (1978).

Martell, A.E. The Design of Chelating Agents, in: Second International Symposium on the Design of Chelating Agents for the Treatment of Cooley's Anemia, Ed. Martell, A.E., Elsevier Publishing Co., Amsterdam, 1981 (in press).

Martell, A.E., The Chemistry of Carcinogenic Metals, Environ. Health Perspect., 1981 (in press).

Martell, A.E. and Smith, R.M., "Critical Stability Constants, Vol. 1, Amino Acids", Plenum Press, Inc., New York, 1974.

Nielsen, F.H., "Newer" Trace Elements in Human Nutrition, Food. Technology, 28, 38-44 (1974).

Nielsen, F.H., Possible Functions and Medical Significance of the Abstruse Trace Metals, in: "Inorganic Chemistry in Biology and Medicine" ACS Symposium Series 140, Ed., Martell, A.E., American Chemical Society, Washington, D.C., 1980, pp. 23-42.

Perrin, D.D., "Stability Constants of Metal-Ion Complexes Part B - Organic Ligands", Pergamon Press, London, 1979.

Rubin, M.; Gignac, S.; Bessman, S.P.; and Belknap, E.L., Enhancement of Lead Excretion in Humans by Disodium Calcium EDTA, Science, 117, 659-660 (1953).

Rubin, B.; Gohil, R.; Martell, A.E.; Motekaitis, R.J.; Penhos, J.C.; and Weiss, p., Computer-Predicted Drugs for Urolithiasis Therapy, in: "Inorganic Chemistry in Biology and Medicine", ACS Symposium Series 140, Ed., Martell, A.E., American Chemical Society, Washington,D.C., 1980, pp. 75-90.

Rubin, M. and Martell, A.E., The Implications of Trace Metal - Nitrilotriacetic Acid Speciation on its Environmental Impact and Toxicology, Biol. Trace Elem. Res., 2, 1-19 (1980).

Saffer, A. and Toribora, T., Changes in Serum and Spinal Fluid Calcium Effected by Disodium Ethylenediaminetetraacetate, J. Lab. Clin. Med., 58, 452-547 (1961)

Schwarz, K., New Essential Trace Elements: Progress Report and Outlook, in: "Trace Element Metabolism in Animals - 2", Eds. Hoekstra, W.G.; Suthe, J.W.; Ganther, H.E.; and Mertz, W., University Park Press, Baltimore, MD, 1974, pp. 355-380.

Sillen, L.G. and Martell, A.E., "Stability Constants of Metal-Ion Complexes", Special Publication No. 17, The Chemical Society, London, 1964.

Sillen, L.G., and Martell, A.E., "Stability Constants of Metal-Ion Complexes" Supplement No. 1, Special Publication No. 25, The Chemical Society. London, 1971.

Smith, R.M. and Martell, A.E., "Critical Stability Constants, Vol. 2, Amines" Plenum Press, Inc., New York, 1975.

Smith, R.M. and Martell, A.E., "Critical Stability Constants, Vol. 4, Inorganic Complexes", Plenum Press, Inc., New York, 1977.

De Simone, R.E.; Penley, M.W.; Charbonneau, L.; Smith, S.; Wood, J.M.; Hill, H.A.O.; Ridsdale, P.1 and Williams, R.J.P., The Kinetics and Mechanism of Cobalamin-Dependent Methyl and Ethyl Transfer to Mercuric Ion, Biochem. Biophys. Acta, 304, 851-863 (1973).

Wood, J.M. and Brown, D.G., The Chemistry of Vitamin B_{12}-Enzymes, Structure and Bonding, 11, 47-105 (1972).

Wood, J.M., Biological Cycles for Toxic Elements in the Environment, Science 183, 1049-1052 (1974).

ENVIRONMENTAL FATE PROCESS 3: HYDROLYSIS IN AQUATIC SYSTEMS

Mel Suffet and Judi Harris

Background

Hydrolysis is likely to represent a significant transformation pathway for some organic chemicals in macro or micro environments where bulk water is present, including: natural surface waters, sediments, interstitial soil water, and ground water. On the other hand, hydrolysis is unlikely to be important as a fate process for organic chemicals in the atmospheric environment or other media where solvent water is absent. Hydrolysis has generally not been considered as an important pathway for transformation of inorganic chemicals, although it can affect the speciation of metals in solution. In hydrolysis reactions the reactant generally reacts with water to form products having the elements of water

$$RX + H_2O \qquad ROH + HX$$

Mechanisms include unimolecular cleavage of bonds followed by addition of water, bimolecular nucleophilic attack by water or hydroxide ion, and acid catalyzed necleophilic attack by water. Catalysis by a variety of naturally occurring metal ions, anions, clays, suspended solids, and organic materials is also possible (Wolfe, 1980). The actual rate of a specific hydrolysis reaction depends on the rate constants for each discrete molecular mechanism and the concentrations of reactant such as H^+ or OH^-; thus pH and temperature are important controlling variables.

It is generally accepted that the hydrolysis rate expression must be applicable over the pH range 5-9 and at temperatures of 5-30°C to be useful for estimates in most natural waters. (Baughman and Burns, 1980; Gomma et. al., 1972; Mill et. al., 1980).

The hydrolysis rate law can be expressed as a function of pH by separating the overall rate expression into its component acid, base and neutral processes:

$$\text{Rate} = d[C]/dt = (k_A[H_3O^+] + k_B[OH^-] + k_N)[C] \qquad 3.2$$

$$k_h = h_A [H_3O+] + k_B[OH^-] + k_N \qquad 3.3$$

$$\ln(C_t/C_c) = k_h t \qquad 3.4$$

The rate constants k_A, k_B, k_N can be estimated in some cases by analogy with closely related chemials or from data measured at higher temperatures by extrapolating to 25° using the Arrhenius equation; rates measured in mixed solvents sometimes can be extrapolated to pure water if the sensitivity of the reaction to solvent is not great or is known with some precision (Mabey and Mill, 1978). Dilling (1975) and Mabey and Mill (1978) have compiled rate constants for hydrolysis for a wide variety of chemical classes in water at or near 25°.

An important and unsolved problem is how to decide if a hydrolysis reaction is important in a specific environmental setting. Generally if the rate of hydrolysis is so fast that over 75% of the chemical has hydrolyzed in an hour or less, one may assume that hydrolysis will be a dominant process; however if the hydrolysis reaction is very slow (half-life over a year) the decision as to its importance must await outcome of similar studies of other loss processes such as photolysis or biotransformation. For example, in the case of the priority pollutants bis-2-chloro ethers, hydrolysis is indicated as one of the fastest mechanisms of transformation, despite the fact that its half life is estimated to be 0.5-2 years (Callahan et. al., 1980). Furthermore, many chlorinated aliphatics that are not sorbed strongly on soil, and are hard to degrade by microorganisms could persist in ground water where hydrolysis may be the only remaining transformation mechanism. In these cases even very slow processes can be important.

Except in rare cases, the effects of environmental catalysts other than H^+ or OH^- are poorly understood. Generally, hydrolysis rates in natural waters and distilled water have shown no dramatic differences, but a few discrepancies have been encountreed (Wolfe, 1980a). A catalytic effect of humic materials on the hydrolysis of pesticides is reported (Khan, 1978; Guy et. al., 1980; Liand Felbeck, 1972). On the other hand, humic materials have recently been shown to retard the hydrolysis of n-octyl esters of 2,4-D (Wolfe, 1980b). Also, the effects of suspended solids and bottom sediments is expected to decrease or stop the rate of hydrolysis (Mill, 1980). Yet, in a recent study, suspended solids have not been shown to have an effect on the hydrolysis of benzyl chloride (Wolfe, 1980c).

Critique of SARs for Estimation of Hydrolysis Rate Constants

Investigation of the rates and mechanisms of hydrolysis reactions has been a major research area in physical organic chemistry for several decades and the detailed mechanisms that are operative in mixed aqueous-organic solvent systems are reasonably well understood for a variety of organic substrates. SARs in the form of linear free energy relationships (LFErs) have been developed for several categories of compounds including carboxylic acid esters, amides, acyl halides, benzyl halides, phosphate esters, carbamates, and others (Chapman and Shorter, 1972; 1978; Lessler and Grunwald, 1963; Wells, 1968). (In recent years attention hands focused on determination of rate constant and development of LFERs for hydrolysis in purely aqueous solvent, to generate relationships with more predictive value for aquatic environments).

Qualitative SARs are useful estimators of relative rates of hydrolysis under similar conditions. Thus leaving groups can be ranked to reflect the relative ease of hydrolyzing specific reactions centers: $X > OC(O)R > OR > NR_{22}$.

The principal quantitative SARs that are presently available for estimation of hydrolysis rate constants are: The Hammett Equation, the Taft Equation, and the Bronsted-type empirical correlations of k_h with pKa of the leaving group. Table 1 presents a brief summary of these correlations, examples of applications, strengths, weaknesses, and current status. A number of LFERs have recently been developed for hydrolysis reactions under conditions corresponding to aquatic environments -- pH 5-9; T 5-30, pure water solvent (Mill et. al., 1980). From the available data, it can be inferred that LFERs of the Hammett, Taft, or Bronsted type could be developed for hydrolysis of many other sets of compounds, each set prescribed by a sufficiently restricted range of structural variation. Relatively high accuracy can be expected for interpolation to other specific compounds whose structure places them within what Hansch refers to as the "explored parameter space," (Hansch, 1981).

It should be noted, however, that the reported value of the correlation coefficient, r, for LFERs based on few data points (e.g., <8) is generally not a reliable measure of precision of the correlation has been suggested (Mill et al., 1980; Exner, 1963). Extrapolation of correlations to compounds or substitutents outside of the range may be more risky in the case of hydrolysis rates than in some other correlations because of the possibility of fundamental changes in hydrolysis reaction mechanism with significant change in structure.

A minimum condition for confirmation of the constancy of mechanism is that formation of analogous hydrolyis product(s) from each reaction in a series used to develop an LFER should be demonstrated. Determination of reaction product distribution has, unfortunately, not always been part of earlier experimental programs. More subtle shifts in reaction mechanism, which are not reflected in a change in product distribution, can also occur.

LFERs with broad applicability seem unlikely to be developed for hydrolysis reactions. Even for such similar series of reactions as hydrolytic displacement of substituted phenoxide ion from phenyl carbamates vs phenyl n-methyl carbomates, Wolfe et. al., (1978) found significantly different regression lines in plots of k_h vs pK_a of the leaving group.

Research and Development Needs

The following are suggested as priority areas for further R&D that could be expected to significantly advance the state-of-the-art of SARs for hydrolysis reaction rate constants in the near-to-mid term.

- Expand and improve the data base of rate constants for hydrolysis in pure water that are reasonably representative of the aquatic environment affecting k.
- Examine the applicability of SARs based on reactions in organic solvents.

- Develop LFERs for additional reaction series using k's for realistic environmental conditions
- Develop explicit set(s) of criteria for identifying and predicting probable mechanistic changes.
- Explore present areas of uncertainty in environmental hydrolysis mechanisms, with particular attention to potential secondary effects of metals, clays, sediments, natural organics, etc.
- Investigate hydrolysis reaction products; develop criteria for predicting formation of products that may be of equal or greater concern than parent. (Perhaps this can be done by organic compound category?)
- Explore the potential utility of other molecular descriptors of R and X that may lead to new and more useful SARs for hydrolysis. For example, one might consider the different properties of leaving groups such as their polarizability or pKs.

TABLE 1. LFERs FOR HYDROLYSIS REACTIONS

1. Hammett Equation

Molecular Descriptor: Substituent constant, σ, for substituents at m- or p- position or aromatic ring in related series of compounds

SAR Equation: $\log k = \rho\sigma + \log k_o$

where: k = rate constant for compound of interest
ρ = reaction constant; slope of LFER for series of related compounds
k_o = rate constant for compound with $\sigma = o$

Example: Hydrolysis of benzyl bromides, methyl benzoates and phenyl acetates (Jaffe, 1953; Mishioka et al., 1975; Wells, 1968)

Strengths/Advantages: Large data base of σ values. Good correlations ($r > 0.99$) for range of hydrolysis rections (mostely in mixed organic-aqueous solvents.

Weaknesses/Disadvantages: ρ values are available only for a limited numer of hydrolysis reaction series.

Many literature studies are for nonaqueous or mixed solvents and not directly applicable to natural waters. Most studies for pure water systems include only 3-5 data points.

There are no Hammett σ values for ortho substitutents because of the probable steric effect complications.

Current Status: Some published applications for reactions in dilute aqueous solution. Potentially applicable to wide range of compounds if ρ value data base is expanded.

Has not yet been validated by measurement of hydrolysis in pure water and comparison to predicted value.

2. Taft Equation

Molecular Descriptors: Polar substitutent constant, σ^*, and steric substituent constant, E_s, for substituents in alkyl system in related series of compounds.

SAR Equation: $\log k = \rho^* \sigma^* + \delta E_s + \log k_o$

where: k = rate constant for compound of interest

ρ^* = reaction constant, measure of sensitivity to polar effects of substitents

δ = reaction constant, measure of sensitivity to steric effects of substitutents

k_o = rate constant for compound with $\sigma^* = E_s = o$.

Examples of Application: Hydrolysis of phthalic acid esters of substituted methanols (Wolfe et al., 1980)

Strengths/Advantages: Moderately large data base of σ^* values.

Weaknesses/Disadvantages Limited data base of E_s values. Limited data base of ρ * and σ^*values. Most published correlations based on 3-6 data points. Two variable equations may give spuriously good fit to limited data set and led to inaccurate extrapolations. Polar/steric effects are more complex and interactive in aliphatic (Taft) systems than in aromatic (Hammett).

Current Status: A few good sets of published data (e.g., Wolfe et al. 1978 carbamate studies). No data for hydrolysis of simple aliphatics (alkyl halides, etc.) Has not yet been validated by measurement of hydrolysis rate and comparison to predicted value.

3. Brönste Approach (Correlation with pk_a of leaving Group)

Molecular Descriptor: Acid dissociation constant (pk_a) for conjugate acid of leaving group in hydrolytic displacement reaction.

SAR Equation: $\log k = m\ pk_a + b$

where: k = rate constant for hydrolysis of compound of interest m,b = slope and intercept of correlation equation

Examples of Applications, hydrolysis of triphenylphosphates and N-methyl phenyl carbamates (Wolfe et. al., 1978; 1980e)

Strengths/Advantages: pk_a values are generally available or can be estimated for leaving groups.

Weaknesses/Diadvantages: Applies only to hydrolytic displacement of weak acid leaving groups (not halide etc.)

Current Status: A few good sets of published data (Wolfe et. al., 1978, 1980d)

REFERENCES

Baughman, G.L., and Burns, L.A., "Transport and Transformation of Chemicals in the Environment: A Perspective in Handbook of Environmental Chemistry," O. Hutzinger, editor. Springer-Verlag; Amsterdam, The Netherlands, 1980.

Mill, T., Mabey, W.R. and Hendry, D.G., "Hydrolysis in Water" - Lab Protocols for Evaluating the Fate of Organic Chemials in Air and Water-T. Mills, W. R. Mabey et al., Contract #68-03-2227, Project Officer, Jim Falco, EPA 560/11-79-012, pp. 5-47, 1980.

Wolfe, N.L., "Determining the Role of Hydrolysis in the Fate of Organics in Natural Waters" in Dyanmics, Exposure and Hazard Assessment of Toxic Chemicals, Haque, R., editor. Ann Arbor Science, 1980a.

Gomma, H.M., Suffet, I.H. and Faust, S.D., "Kinetics of Hydrolysis of Diazinon and Diazoxon, Residue Reviews, 29, 171, 1972.

Callahan et. al., "Water-Related Environmental Fate of 129 Priority Pollutants', Vol. 1, EPA 440/4-79-029a, Vol. 2, EPA 440/4-79-403, 1980d

Wolfe, N.L., "Organophosphate and Organophosphorothionate Esters: Applications of Linear Free Energy Relationships to Estimate Hydrolysis Rate Constants for Use in Environmental Fate Assessment, Chemosphere, 9, 571, 1980e.

Wolfe, N.L., Zepp, R.G. and Paris, D.F., "Use of Structure-Reactivity Relationships to Estimate Hydrolytic Persistence of Carbamate Pesticides", Water Research, 12, 561, 1978.

Mabey, W. and Mill, T., "Critical Review of Hydrolysis of Organic Compounds in Water Under Environmental Conditions" J. Phys. Chem. Ref. Data, 7, 383, 1978.

Li, G.C. and Felbeck, J.T., Jr., "Atrazine Hydrolysis at Catalyzed by Humic Acids", Soil Science, 114, 201, 1972.

Kahan, S.U., "Kinetics of Hydrolysis of Fulvic Acid Solution", Pesticide Science 9, 39, 1978.

Guy, R.D., Narine, D.R. and DeSilva, S., "Organocation Speciation. I. A Comparison of the Interations of Methylene Blue and Paraquat with Bentonite and Humic Acid", Can. J. Chem., 58, 547-554 (1980)

Lessler, J.E. and Grunwald, E., "Rates and Equilibria of Organic Reactions", John Wiley and Son, New York, 1963.

Wells, P.R., "Linear Free Energy Relationships", Academic Press, London, 1968.

Chapman, N.B. and Shorter, J., Eds., Advances in Linear Free Energy Relationships, Plenum Press, New York, 1972.

Chapman, N.B., and Shorter, J., Eds., Correlation Analyis in Chemistry - Recent Advances, Plenum Press, New York, 1978.

Exner, O., "The Hammett Equation - The Present Position. Chapter 1 in J.E. Lessler and E. Grunwald, "Rates and Equilibria of Organic Reactions", John Wiley and Sons, New York, 1963.

Dilling, W.L., Tefertiller, N.B., and Kallos, G.J., (1975), "Evaporation Rates of Methylene Chloride, Chloroform, 1,1,1-Trichloroethane, Trichloroethylene, Tetrachloroethylene and Other Chlorinated Compounds in Dilute Aqueous Solutions, Environmental Science & Technology, 9, 883.

Wolfe, N.L., (1980b), "Catalyzed Hydrolysis Studies" - U.S. EPA Newsletter, Quarterly Research Highlights - Environmental Research Labs, Athens, GA, Volume 4, #4, p. 4, Sept. - Dec. 1980.

Hansch, C., (1981), "Plenary Address", SAR's Workshop, San Antonio, TX.

Mill, T., (1980), "Data Needed to Predict the Environmental Fate of Organic Chemicals", in R. Hawue, Ed., Dynamics, Exposure and Hazard Assessment of Toxic Chemicals, Ann Arbor Science

Jaffe, H.H., (1953), "A re-examination of the Hammett Equation", Chemical Review, Vol. 53, pp. 191-261.

Mishioka, T., Fujita, T., Kitamura, K., and Makajama, M., "The Ortho Effect in Hydrolysis of Phenol Esters", J. of Org. Chem. 40, pp. 2520-2525(1975).

ENVIRONMENTAL FATE PROCESS 4: OXIDATION IN WATER by T. Mill

Background

Oxidation is a major loss process in the environment where significant concentrations of oxidants are generated by photochemical processes involving natural and anthropogenic light absorbers. In urban atmospheres, both HO• radical and O_3 are generated by a complex photocycle involving NO_2, O_2, and organic pollutants. In natural water, RO_2• radical and singlet O_2 ($^1O_2$) are generated by photolysis of humic materials dissolved or suspended in water.

Direct photooxidation of many molecules also occurs by way of excited state interactions with triplet O_2, but the rates of these processes are controlled by the photophysics and hotochemistry of excited states, not by the thermal chemistry of oxidant-molecule interaction (see Photochemistry section).

Recent reviews of oxidation in water and in air (Mill et al., 1979; Hendry and Kenley, 1979) suggest that only a few selected oxidants are probably important in water. Table 1 lists these oxidants and their average diurnal concentrations.

TABLE 1. OXIDANT CONCENTRATIONS IN WATER AND AIR

Oxidant	Concentration, M
	Water[a]
RO_2	1×10^{-9}
$^1O_2$	1×10^{-12}

(a) From Mill (1979)

SAR for Oxidation

The concentrations and known reactivities of the major oxidants provide a simple basis on whch to separate reactive and nonreactive classes of chemicals in oxidations in water. Those classes of chemicals that have half-lives of more than one year may be considered as inert to these chemical oxidants and thus can be excluded from screening tests for oxidation. Tables 2 and 3 list half-lives for a variety of organic structures in reactions with RO_2 and $^1O_2$ based on the reasonable assumption that the rate of a specific oxidation process follows the relation

$$-d[C]/dt = k_{OX}[OX][C]$$

If [OX] is constant* equation (21) becomes

$$d[C]/dt = k_{OX}[C]$$

and the halflife for chemical C is

$$t_{1/2} = \ell n2/k_{OX}$$

Oxidation by RO_2

The range of measurable reactivities of organic compounds towards RO_2 radicals covers about 10^{10} in k_{OX} (Howard, 1972). Only chemicals that react with rate constants $>10^3$ M^{-1} s will be

---

*This assumption is usually referred to as a steady-state assumption, and values used for [OX] take into account the fact that oxidants depend on sunlight and thus change in concentration diurnally.

TABLE 2. RATES OF OXIDATION BY $RO_2\cdot$ RADICAL IN SOLVENTS AT 30°C

Class	k_p^a $M^{-1}\ s^{-1}$	Half-Life, $t_{1/2}^b$ days
Hydrocarbons	0.01	8×10^5
Olefin	0.09	9×10^4
Benzyl	1	8×10^3
Aldehyde	0.1	8×10^4
Alcohol	0.01	8×10^5
Phenol	1×10^4	0.8
Aromatic amine	1×10^4	0.8
Hydroquinone	1×10^6	12 min
Hydroxylamine	1×10^5	120 min
Hydroperoxide	1×10^5	120 min
Polycyclic aromatic	1×10^3	8

[a] Per reactive X-H bond : from Hendry et al. (1974) and Howard (1972).

[b] For 1×10^{-9} M RO_2; $t_{1/2} = \ln 2/8.64 \times 10^{-5}$ kp in days.

TABLE 3. RATES OF OXIDATION BY SINGLET OXYGEN IN SOLVENTS AT 25°C

Class	$k_{^1O_2}$[a] $M^{-1}\ s^{-1}$	Half-Life[b], $t_{1/2}$
Alkanes	$< 2 \times 10^2$	> 100 yrs
Alcohols and ethers	$< 2 \times 10^2$	> 100 yrs
Ketones and Aldehydes	$< 2 \times 10^2$	> 100 yrs
Aromatics (including amines and phenols)	$< 2 \times 10^2$	> 100 yrs
Alkyl halides	$< 2 \times 10^2$	> 100 yrs
Acids and esters	$< 2 \times 10^2$	> 100 yrs
Unsubstituted olefin	3×10^3	7.3 yrs
Cyclic olefins	2×10^5	40 days
Substituted olefin	1×10^6	8.0 days
Dialkyl sulfide	7×10^6	27 hrs
Diene	1×10^7	19 hrs
Imidazoles	4×10^7	4.8 hrs
Furans	1.4×10^8	1.0 hrs
Trialkyleneamines	8×10^8	14 mins

[a] Per molecule from Foote (197) and Mill et al. (1979).

[b] For 1×10^{-12} M $^1O_2$; $t_{1/2} = \ell n\ 2/k_{^1O_2} \times 10^{-12}$ in sec.

oxidized in aquatic systems at significant rates (Table 2). Therefore, we can exclude from testing any compound that has only CH or aromatic structures and all of their simple derivatives including carbonyls, esters, alcohols, halides, aliphatic amines, and ethers.

Aromatic amines and phensols are two major classes of chemicals that are reactive enough toward $RO_2\bullet$ radical to warrant detailed consideration regarding SRC. Fortunately, because these compounds are useful as antioxidants, several investigators have measured absolute rate constants for H-atom transfer to RO_2 from many types of phenols and some amines (Howard and Ingold, 1963; Howard 1972).

$$RO_2\bullet + XPhOH \xrightarrow{k_{inh}} RO_2H + XPhO$$

Table 4 summarizes the absolute rate constants (k_{inh}) for selected classes of phenols; values of ρ^+ for each subseries of phenols are also listed. In all cases ρ^+ is negative, indicating that the transition state has some postive character stabilized by electron donor substituents. The reactivity of all measured phenols ranges from a low of about $3 \times 10^2\ M^{-1}\ s^{-1}$ for 4-CNPhOH to a high of about $9 \times 10^4\ M^{-1}\ s^{-1}$ for k_{inh} for hydroquinone extrapolated to the same temperature. Aromatic amines have been subject to less systematic study than phenols; nonetheless, the data listed in Table 5 shows that as a group, amines are as reactive or more so than phenols: values of k_{inh} range from a low of $9 \times 10^2\ M^{-1}s^{-1}$ for $MeO_2CPhNHMe$ to highs of $2 \times 10^5\ M^{-1}\ s^{-1}$ for

TABLE 4. SAR FOR OXIDATION OF PHENOLS BY $RO_2\cdot$ RADICAL IN SOLVENTS

Substituent X	$10^{-4}\ k_{inh}\ M^{-1}\ s^{-1}$
XPhOH at 65°C with polystyrylperoxy radical[a]	
4-HO	9.3
4-Me	0.92
H	0.23
3-Cl	0.056
4-CN	0.030
$\rho^+ = -1.49$ in styrene	
$\rho^+ = -3.7$ in chlorobenzene	
$\rho^+ = -3.48 \mp 0.114$ in methyl methacrylate	
$2,6\text{-}(t\text{-}Bu)_2\text{-}4\text{-}X\text{-}PhOH$ at -37°C with $t\text{-}BuO_2\cdot$[b]	
MeO	2.3
Me	0.50
H	0.34
Cl	0.50
CN	0.10
$\rho^+ = -1.00$, $r = 0.94$ in isopentane	
o-Alkyl PhOH at 65°C with polystyryl peroxy radical[c]	
$2,6\text{-}Me_2\text{-}4\text{-}XPhOH$	
Me	2.8
Cl	1.2
CN	0.73
H	0.89
$\rho^+ = -1.36 \mp -1.46 \pm 0.115$	

TABLE 4. (Concluded)

Substituent X	10^{-4} k_{inh} M^{-1} s^{-1}
2-t-Bu-n-XPhOH[c]	
H	0.75
4-MeO	10.7
5-Me	1.1
4-IN	0.087
	$\rho^+ = -1.46 \pm 0.012$
2,6-(t-Bu)$_2$-4-XPhOH[c]	
H	0.23
MeO	3.1
Me	0.80
Cl	0.32
CN	0.077
NO$_2$	∿ 0.05
	$\rho^+ = -1.11 \pm 0.026$

[a] Howard and Ingold, 1963a.
[b] Howard and Furimsky, 1973.
[c] Howard and Ingold, 1963b.

TABLE 5. SAR FOR OXIDATION OF AMINES BY $RO_2\cdot$ RADICAL IN SOLVENTS

Amine	$10^{-4}\ k_{inh}\ M^{-1}\ s^{-1}$
N-Methylanilines at 65°[a,b]	
4-H	0.4
4-Me	1.2
3-Me	0.5
4-MeOC(O)	0.09
4-MeO	1.9
$\rho = -1.6$	
Diphenylamines at 65°[a,b]	
4,4′-H_2	4
4,4′-Me_2	10
4,4′-$(MeO)_2$	33
4,4′-$(NO_2)_2$	0.16
4-MeO	20
3-Cl	1.8
4-NO_2	0.6
$\rho^+ = -0.89$	
Napthylamines at 40°C[a,b]	
N-α-Napthyl	0.15
N-α-Naphthyl-N-Phenyl	1.25
N-β-Napthyl-N-Phenyl	10
N,N-Di-β-Napthyl	18

a In reaction with poly(peroxystyryl)peroxy radical.

b From Brownlie and Ingold (1967).

diphenylamines. Values of ρ are negative and are similar to those for phenols (-0.8 to 0.6).

All these data were measured in nonpolar organic solvents. No useful data are available to indicate the extent of solvent effects on the process, for use in extrapolating to water; however, Howard and Ingold (1963) have speculated that ρ may increase with an increase in the solvent polarizability. For the present we probably can use the relative values of k_{inh} for phenols and amines with few reservations, but the absolute values in water may prove to be different from those in organic solvents.

Oxidation by Singlet Oxygen ($^1O_2$)

Zepp et. al., (1978) demonstrated that sunlight irradiation of natural waters led to production of $^1O_2$ at average concentrations of 1×10^{-12}M. This value when combined with rate constants for oxidation of a variety of organic structures (Table 3) lead to predictions of very long half-lives for most simple structures. Reactive molecules that warrant laboratory testing are electron-rich molecules such as branched olefins, enamines, polycyclic aromatics. and sulfides.

195

Estimation Methods

SAR for $^1O_2$ oxidations are very limited. Foote and Denny (1971) measured the relative rates of oxidation of a series of substituted styrenes and found that for allylic oxidation correlation by the Hammett equation (σ) was satisfactory and $\rho = -0.92$. However, epoxidation of the styrene double bond which accompanies allylic oxidationis not correlated well by σ nor σ^+, but is correlated by $[\sigma - 0.37\ (\sigma^+ - \sigma)]$ to give $\rho = -0.87$.

Research Needs

SAR for oxidations in water are virtually nonexistent; as a result one must extrapolate from reactivity data in non polar solvents with little reliable information about the effects of solvents on these reactions. Some fundamental studies are needed of the reactions of both singlet oxygen and peroxy radicals with organic chemicals in water including measurements of absolute rate contants for the important (rapid) oxidations.

REFERENCES

Brownlie, I.T., and K.U. Ingold. 1967. The Inhibited Autoxidation of Styrene. Part VI. The Relative Efficiencies and the Kinetics for Inhibition by N-aryl Anilines and N-alkyl Anilines. Can. J. Chem. 45:2419

Foote, C.S. 1976. Free Radicals in Biology. W.A. Pryor, ed. Academic Press, Inc., New York.

Foote, C.S., and R.W. Denny. 1971. Chemistry of Singlet Oxygen. XIII. Solvent Effects on the Reaction with Olefins. J.Am. Chem. Soc., 93:5168.

Hendry, D.G., T. Mill, L. Piszkiewicz, J.A. Howar, and H.K. Eigenmann. 1974. A Critical Review of H-Atom Transfer in the Liquid Phase. Chlorine Atom, Alkyl, Trichloromethyl, Alkoxy and Alkylperoxy Radicals. J. Phys. Chem. Ref. Data 3:937.

Howard, J.A., and K.U. Ingold. 1963. The Inhibited Autoxidation of Styrene. Part II. Can. J. Chem. 41:1744.

Howard J.A., 1972. Absolute Rate Constants for Reactions of Oxyl Radicals. Adv. Free Radical Chem. 4:49.

Mill, T. 1979. Data Needed to Predict Environmental Fate of Organic Compounds. Proceedings of Symposium on Environmental Fate held at meeting of Amer. Chem. Soc., Miami, FL, Sept. 1978. Ann Arbor Publishers, in press.

Zepp, R.G., N.L. Wolfe, G.L. Baughman, and R.C. Hollis. 1978. Singlet Oxygen in National Water. Nature 278:421.

ENVIRONMENTAL FATE PROCESS 5: BIOTRANSFORMATION PROCESSES by Perry L. McCarty

Background

Only a few decades ago the concept of microbial infallibility, that all organic molecules were biodegradable to inorganic end products, was generally accepted. Mechanisms and pathways for enzymatic hydrolysis, fermentation, and oxidation of the then naturally occurring organic materials were (or are now) relatively well known. Thus, for naturally occurring materials, the relationship between structure and activity for the biochemical processes in general is quite well established. These same rules can also be applied to synthetic organic chemicals, except for those for which the structures are significantly different from natural organic materials. The task then is to evaluate structural differences between natural and synthetic organic chemicals, and to evaluate how these structural differences may affect biotransformations. In investigating this problem, it should be recognized that many natural organic materials are also quite resistant to biological alteration. Examples are lignin, which consists of large molecules with aromatic structures connected with carbon-carbon or ether linkages that resist enzymatic hydrolysis, and steroids that contain quarternary carbon linkages. Knowledge about why such compounds resist biotransformation can also be applied to understanding structural features of synthetic chemicals that make them resist biotransformation.

Biotransformation is used here to mean any alteration that changes one compound to another and is mediated by enzymes that are either free or associated with an organism. Biochemical reactions can also change the chemical environment in ways that either increase or decrease the rate of a chemical reaction, e.g., lowering pH through fermentation of carbohydrates to produce acids, or changing pE by reduction of sulfate to sulfide. The effect of the latter on rates, however, is not referred to here as resulting from "biotransformation." Biotransformation can result in hydrolysis, oxidation, reduction, elimination, addition, and substitution. These processes together can result either in a simple alteration of an organic molecule or in its mineralization which is complete oxidation to inorganic end products.

A fair amount of literature exists suggesting general structural features that affect biotransformations (Leinger, 1979; National Academy of Sciences 1972; Kaufman and Planner, 1972, Alexander, 1981). Much of this is related to pesticides, but the concepts are useful for understanidng other molecules as well. For the most part, these generalizations result from empirical observations that are not correlated to molecular descriptors to the extent they have some other properties affecting fate in the environment.

The relative rate of different biotransformation processes is not readily measured, especially when dealing with a transformation by whole organisms rather than specific enzymes. Ideally, relative biotransformability would be reported in terms of rate of transformation per unit mass of transforming organism or enzyme per unit time under given chemical conditions and temperature. However, it is generally measured in terms of the time required to develop an acclimated mixed culture capable of bringing about a given transformation. Acclimation here may result from organism selection and/or enzyme induction. Acclimation times of one to two months have been observed for certain compounds, especially under an aerobic conditions. A relationship does not necessarily exist between time to acclimate and rate of transformation within acclimated culture. Therefore, one must be cautious in interpreting the available data. This area needs better definition.

Another area of concern is the particular group of organisms carrying out the transformation. Bacteria and fungi as classes perhaps have the greatest diversity of abilities in this regard, and since they are the most universal decomposers in the environment, attention should be focused upon them. However, because of data availability, most generalizations below are derived from studies with bacteria rather than fungi.

Biotransformations

The major biochemical reactions brought about by heterotrophic microorganisms are hydrolysis, oxidation, reduction, addition, elimination, and substitution. Hydrolysis of esters, glycosidic bonds, peptide and other C-N bonds, and acid anhydrides occurs by the removal of two hydrogens from adjacent atoms and addition of water across the resulting double bind, although direct addition of molecular oxygen also occurs, such as in the initial stages of some aromatic oxidations. The hydrogens (and associated electrons) so removed are transferred to appropriate terminal electron acceptors, oxygen being the preferred acceptor under aerobic conditions, and nitrate, sulfate, and carbon dioxide, in that order, under anaerobic conditions. Organic materials may also act as electron acceptors, internal oxidation and reduction of a molecule taking place during fermentation. By such a process, an unsaturated compound in the environment may be reduced. Oxidation can result in the complete mineralization of an organic compound, except in feremcontation or when carbon dioxide is the terminal electron acceptor. In the latter case, methane is an end product.

It is commonly believed that transformations are less likely under anaerobic than under aerobic conditons. However, there are important exceptions. For example, the pesticides lindane and DDT, and the brominated and chlorinated trihalomethanes are readily transformed under anaerobic conditions but are quite persistent under aerobic conditions. Thus, even here, one needs to be cautious about generalizations. The possibilities for transformation, however, are very dependent upon the chemical environment.

In a recent review, Alexander (Hill and McCarty, 1967; Borwer et. al., 1981) listed type biological reactions for transformation (Table 1) and cleavage (Table 2) of organic chemicals that have been reported, and gave extensive references for each. These listings indicate the wide variety of biotransformations that are possible. Alexander also emphasized the importance intermediates of biotransformation may have in the environment, the present difficulties in making generalizations about which intermediates are likely to accumulate, and the inability to predict in general which steps in the process will be slow and which fast. Thus, the present ability to making meaningful . . . "predictions about which intermediates will appear outside the cell and accumulate in the environment are tenuous at best."

Reaction Kinetics

Few generalizations can be made about rates for biochemical reactions, except perhaps for the maximum rates for substrate oxidations that yield energy to the organism. With bacteria, the maximum rate of oxidation for many organic compounds appears to be limited primarily by electron transport rather than by the structure of the molecule being degraded. These maximum rates may be used for comparative purposes. For bacteria this maximum rate (Rmax) is about one electron-mole per day per gram of bacteria at 20°C (5). As with chemical reactions, this rate about doubles with each 10° C rise in

Table 1. Type reactions for transformation of chemicals of environmental importance. (a)

Category	Reaction*	Example†
Dehalogenation	$RCH_2Cl \rightarrow RCH_2OH$	Propachlor (S,M)
	$ArCl \rightarrow ArOH$	Nitrofen (S)
	$ArF \rightarrow ArOH$	Flamprop-methyl (S)
	$ArCl \rightarrow ArH$	Pentachlorophenol (S,M)
	$Ar_2CHCH_2Cl \rightarrow Ar_2C{=}CH_2$	DDT (M)
	$Ar_2CHCHCl_2 \rightarrow Ar_2C{=}CHCl$	DDT (G,W,M)
	$Ar_2CHCCl_3 \rightarrow Ar_2CHCHCl_2$	DDT (G,S,M)
	$Ar_2CHCCl_3 \rightarrow Ar_2C{=}CCl_2$	DDT (W,S,M)
	$RCCl_3 \rightarrow RCOOH$	N-Serve (S), DDT (W,M)
	$HetCl \rightarrow HetOH$	Cyanazine (S)
Deamination	$ArNH_2 \rightarrow ArOH$	Fluchloralin (S)
Decarboxylation	$ArCOOH \rightarrow ArH$	Bifenox (S)
	$Ar_2CHCOOH \rightarrow Ar_2CH_2$	DDT (M)
	$RCH(CH_3)COOH \rightarrow RCH_2CH_3$	Dichlorfop-methyl (S)
	$ArN(R)COOH \rightarrow ArN(R)H$	DDOD (S)‡
Methyl oxidation	$RCH_3 \rightarrow RCH_2OH$ and/or $\rightarrow RCHO$ and/or $\rightarrow RCOOH$	Bromacil (S), diisopropylnaphthalene (G), pentachlorobenzyl alcohol (S,M)
Hydroxylation and ketone formation	$ArH \rightarrow ArOH$	Benthiocarb (S), dicamba (W)
	$RCH_2R' \rightarrow RCH(OH)R'$ and/or $\rightarrow RC(O)R'$	Carbofuran (S), DDT (S,W,G)
	$R(R')CHR'' \rightarrow R(R')CHOH(R'')$	Bux insecticide (S)
	$R(R')(R'')CCH_3 \rightarrow R(R')(R'')CCH_2OH$	Denmert (S)
β oxidation	$ArO(CH_2)_nCH_2CH_2COOH \rightarrow ArO(CH_2)_nCOOH$	ω-(2,4-Dichlorophenoxy)-alkanoic acids (S,M)
Epoxide formation	$RCH{=}CHR' \rightarrow R\overbrace{CH{-}CH}^{O}R'$	Heptachlor (S,M)
Nitrogen oxidation	$R(R')NR'' \rightarrow R(R')N(\rightarrow O)R''$	Tridemorph (S)
Sulfur oxidation	$RSR' \rightarrow RS(O)R'$ and/or $\rightarrow RS(O_2)R'$	Aldicarb (S,M)
=S to =O	$(AlkO)_2P(S)R \rightarrow (AlkO)_2P(O)R$	Parathion (S,M)
	$RC(S)R' \rightarrow RC(O)R'$	Ethylenethiourea (S)
Sulfoxide reduction	$RS(O)R' \rightarrow RSR'$	Phorate (S)
Reduction of triple bond	$RC{\equiv}CH \rightarrow RCH{=}CH_2$	Buturon (S,M)
Reduction of double bond	$Ar_2C{=}CH_2 \rightarrow Ar_2CHCH_3$	DDT (W,M)
	$Ar_2C{=}CHCl \rightarrow Ar_2CHCH_2Cl$	DDT (W,M)
Hydration of double bond	$Ar_2C{=}CH_2 \rightarrow Ar_2CHCH_2OH$	DDT (W,M)
Nitro metabolism	$RNO_2 \rightarrow ROH$	Nitrofen (S)
	$RNO_2 \rightarrow RNH_2$	Pentachloronitrobenzene (S,M), Sumithion (S,W,M)
Oxime metabolism	$RCH{=}NOH \rightarrow RC{\equiv}N$	Aldicarb (S,M)
Nitrile-amide metabolism	$RC{\equiv}N \rightarrow RC(O)NH_2$ and/or $\rightarrow RCOOH$	Bromoxynil (S,M), Dichlobenil (S,W)

*Abbreviations: R, organic moiety; Ar, aromatic; Alk, alkyl; Het, heterocycle. †Reaction demonstrated in sewage (G), microbial culture (M), soil (S), or natural waters (W). ‡3-(3′,5′-Dichlorophenyl)-5,5-dimethyloxazolidine-2,4-dione.

Table 2. Type reactions for cleavage of chemicals of environmental importance.

Substrate	Reaction	Example
Ester	$RC(O)OR' \rightarrow RC(O)OH$	Malathion (M), phthalates (W,M)
Ether	$ArOR \rightarrow ArOH$	Chlomethoxynil (S), 2,4-D (S,W,M)
	$ROCH_2R' \rightarrow ROH$	Dichlorfop-methyl (S)
C-N bond	$R(R')NR'' \rightarrow R(R')NH$ and/or $\rightarrow RNH_2$	Alachlor (S,M), trimethylamine (G,M)
	$RN(Alk)_2 \rightarrow RNHAlk$ and/or $\rightarrow RNH_2$	Chlorotoluron (S), trifluralin (S,M)
	$RNHCH(R')R'' \rightarrow RNH_2$	Imugam (S)
	$RNH_2CH_2R' \rightarrow RNH_2$	Glyphosate (S,W)
Peptide, carbamate	$RNHC(O)R' \rightarrow RNH_2$ and/or $HOOCR'$	Benlate (S,M), dimethoate (S)
	$R(R')NC(O)R'' \rightarrow R(R')NH + HOOCR''$	Benzoylprop-ethyl (S)
=NOC(O)R	$RCH{=}NOC(O)R \rightarrow RCH{=}NOH$	Aldicarb (S,M)
C-S bond	$RSR' \rightarrow ROH$ and/or HSR'	Benthiocarb (S), Kitazin P (S,M)
C-Hg bond	$RHgR' \rightarrow RH$ and/or Hg	Ethylmercury (S,M), phenylmercuric acetate (S,M)
C-Sn bond	$R_3SnOH \rightarrow R_2SnO \rightarrow RSnO_2H$	Tricyclohexyltin hydroxide (S)
C-O-P	$(AlkO)_2P(S^*)R \rightarrow AlkO(HO)P(S^*)R$ and/or $\rightarrow (HO)_2P(S^*)R$	Gardona (S), malathion (S,M)
	$ArOP(S^*)(R)R' \rightarrow ArOH$ and/or $HOP(S^*)(R)R'$	Diazinon (S), parathion (S,M)
P-S	$RSP(O)(R')OAlk \rightarrow HOP(O)(R')OAlk$	Hinosan (M), Kitazin P (M)
Sulfate ester	$RCH_2OS(O_2)OH \rightarrow RCH_2OH$ and/or $HOS(O_2)OH$	Sesone (S,M)
S-N	$ArS(O_2)NH_2 \rightarrow ArS(O_2)OH$	Oryzalin (S)
S-S	$RSSR \rightarrow RSH$	Thiram (S,M)

*Sulfur or oxygen.

Reproduced from best available copy.

(a) From Alexander (1981)

temperature. The rate is also a function of organic concentration as indicated in the following:

$$R = R_{max} \left(\frac{S}{K_s + S} \right) \qquad (1)$$

where S is substrate concentration in moles/liter and K_s is a constant that is a function of the substrate, the chemical environment, and the species carrying out the reaction. In general, K lies in the range between 10^{-4} and 10^{-6} moles/liter. The degree of resistance to biooxidation might be related to the extent to which the rate by an acclimated culture is below that given by Equation 1. For other biotransformations, such generalizations are few.

Biotransformation Concentration Thresholds Concentrations of synthetic organic chemicals in natural aquatic systems are generally quite low, in the ng/l to ug/l range. Some evidence suggests that enzyme induction may not occur for substrates when the concentrations are so low (Boething and Alexander (1979). However, some substrates are readily transformed at these low concentrations as long as bacteria are present and supported by primary substrates present at higher concentration (McCarty et. al., 1981, Brown). This area deserves more study as it bears directly on the question of whether biotransformations of given compounds can occur in the environment.

Structure Activity Relationships

The preceding discussion indicates that whether or not a compound may be biotransformed depends upon the species present and environmental conditions as well as on the organic molecule itself. There are several generalizations that can be made about structural features that affect a compound's ability to be biotransformed (Kaufmann and Plimmer, 1972). A few examples are worth mentioning.

Straight-chain aliphatic alcohols, aldehydes, and acids are readily biotransformed and mineralized by aerobic and anaerobic cultures. Saturated alkanes appear to be biotransformed only in aerobic environments. In general, increased chain length tends to increase resistance to biotransformation, as does a substitution of an alkyl or halogen for hydrogen within the chain. The more the latter gropus are added, the greater the resistance to transformation. The position of such a substitution also affects the ease of biotransformation. For aliphatic acids, α-halogen additions have little affect, while substitution at other locations greatly reduces ability to be transformed. In general, substitutions that make removal of two hydrogens on adjacent atoms difficult, increase resistance.

Similar generalizations are possible for aromatic compounds. Substitutions with two or more alkyl, halogen, nitro, or sulfonate groups generally result in increased resistance to transformation. Meta substitution, in particular increases resistance to transformation, while ortho and para substitutions generally have much less effect, but this depends upon the group substituted. An exception to the above is the substitution of chlorine in the ortho position of phenoxyacetate, which inhibits transformation, while substitution in the para position enables the compound to be readily transformed.

The effect of substitutent additions to carbamate and organophosphate pesticides has also been extensively studied, allowing one to draw some general conclusions about influence or structure (Kaufmann and Plimmer, 1972). However, space does not permit going into details here. Polymeric materials, in general, owe their resistance to poor solubility and linkages that are resistant to enzymatic hydrolysis. Here, parallels with natural lignins exist.

In summary, the general pathways for microbial transformation and mineralization of many common organic compounds is well known. Substituents that interfere with these recognized pathways often tend to make the molecules more resistant to transformation. However, few quantitative relationships relating structure and activity have been developed.

Research Needs

1. Development of suitable protocol for evaluating the relative ease of biotransformation for compounds, perhaps including a comparative study of relationships, if any,

between acclimation time and rates of transformation by acclimated cultures.

2. Development of understanding of mechanisms underlying the relative ease of dehalogenation of aliphatic compound under anaerobic as compared with aerobic conditions.
3. Evaluation of existing data to determine why from a biochemical view certain structures resist biotransformation and others do not.
4. Experimentation to determine structure activity relationships for structures of importance but with little empirical data, as determined by item 3 above, to make adequate judgments.
5. Investigations of relationships between organic structure and concentration thresholds required for enzyme induction.

REFERENCES

Lehninger, A.L., Biochemistry. Worth Publishers, Inc., New York, 1970. 833 pp.

National Academy of Scieces. Degradation of Synthetic Organic Molecules in the Biosphere. Washington, D.C., 1972. 350 pp.

Kaufman D.D. and J.R. Plimmer. Approaches to the Synthesis of Soft Pesticides. In: Water Pollution Microbiology. Ed. by R. Mitchell. John Wiley and Sons, Inc., New York, 1972. pp. 173-203.

Alexander, M. Biodegradation of Chemicals of Environmental Concern. Science, 211:132-211, 1981.

McCarty, P.L. Energetics andBacterial Growth. In: Organic Compounds in Aquatic Environments. Ed. by S.D. Faust and J.V. Hunter. Marcel Dekker, Inc., New York, 1971. pp. 495-531.

McCarty, P.L., M. Reinhard, and B.E. Rittmann. Trace organics in Groundwater. Environmental Science & Technology, 15(1):40-51, 1981.

Hill, D.W. and Pl.1 McCarty. Anaerobic Degredation of Selected Chlorinated Hydrocarbon Pesticides. Jour. Water Pollution Control Federation, 39:1259-1277, 1967.

Bouwer, E., B.E. Rittmann, and P.L. McCarty. Anaerobic

Degradation of Halogenated one-and Two-Carbon Organic Compounds. Environmental Science and Technology, in press.

Boething, R.S. and M. Alexander. Effect of Concentration of Organic Chemicals on their Biodegradation by Natural Microbial Communities. Applied and Environmental Microbiology, 37(6): 1211-1216(1979).

10. Bouwer, E., Stanford University Water Quality Control Research Laboratory, unpublished data.

FATE OF CHEMICALS IN THE ATMOSPHERE by D.G. Hendry

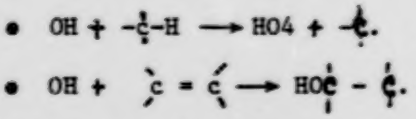
Background

Chemicals emitted into the atmosphere are subject to both physical and chemical processes. We have a reasonably good understanding of these processes and, furthermore, rates of these processes can be readily treated by various predictive methods. In this section we outline the various atmospheric processes and point out that for a large majority of compounds the fate is determined by three chemical processes. The most promising predictive method for each of the three processes is discussed.

In concluding the discussion, the information needed to improve the reliability of these methods is summarized.

The Processes. The chemical atmospheric processes that influences the fate of most chemicals are:

(1) Reaction with OH radical



$$\bullet\ OH + -\overset{|}{\underset{|}{C}}-H \longrightarrow HO4 + -\overset{|}{\underset{|}{C}}\cdot$$

$$\bullet\ OH + \;>C=C<\; \longrightarrow HO\overset{|}{\underset{|}{C}} - \overset{|}{\underset{|}{C}}\cdot$$

OH radicals are present througout the atmosphere during the day. The average concentration is 1×10^{6} radicals cm^{-3} (3×10^{-15}m).

OH is very reactive with most organic compound reacting by hydrogen atom abstraction or addition to carbon-carbon double bonds in olefins or aromatic compounds. The rate of reaction with chemicals is expressed as

$$\frac{d[C]}{dt} = \text{Rate OH} = R_{OH}\ [OH]\ [C]$$

Since [OH] is relatively constant k_{OH} [OH] may be treated as a first-order rate constant and thus the concentrtation of substrate as a function of time can be expressed as:

$$\ell n \left[\frac{C_o}{C_t}\right] = k_{OH}\ [OH]\ t$$

where C_o and C_t are the initial and time-t concentrations of chemical. (2) Reaction with ozone : $>C = C< + O_3 \longrightarrow >C = O + >C = O_2$

The concentration of ozone in the atmosphere averages about 1 x 10^{12} molec cm^{-3} (~ 0.04 ppm) or 2 x 10^{-9}M. The major reaction of ozone is with olefins and highly substituted aromatics. Some compounds containing heteroatoms, especially S and N, may also react, however, data are limited. The rate is expressed

$$\frac{d[O_3]}{dt} = \text{Rate } O_3 = k_{O_3}\ [O_3][C]$$

Assuming O_3 is a constant gives a first-order decay and the following expression is obtained.

$$\ell n \frac{[C]}{[C]} = k_{O_3}\ [O_3]\ t$$

(3) Reaction by Phetolysis

$$C \xrightarrow{hv} \text{Products}$$

Photolysis is limited to those compound that absorb light in the portion of the solar spectrum (290 nm) which is not screened out by stratospheric ozone. The rate may be expressed in the form

$$\text{Rate} = \phi[C] \sum_{\lambda} \sigma \lambda J \lambda$$

where ϕ is th equantum yield, σ_{λ} is the cross section for absorption of light by the compound at wavelength λ, and J_{λ} is the flux of light at wavelength λ.

These chemical processes, while considered to be dominant processes affecting the fate of chemicals, must compete with the following physical processes:

(1) Dilution by Diffusion. Dilution is very fast in the atmosphere but does not affect the loss of mass. Generally emissions are diluted 10^3 fold in the first 20-30 minutes. After an hour diuution is of the order of 10^6 fold. This rapid decrease in concentration reduces the importance of biological effects at downwind loations from the source.

(2) Rainout and Washout. These two processes involve similar basic factors and therefore are considered together. They are important only for compounds that are highly water soluble and then only when it rains. Under conditions where rain is minimal the processes are unimportant.

(3) Dry Deposition. This process is important for chemicals that absorb or react readily at the earth's surface. It generally occurs on a slow time scale and is not a dominant process.

(4) Sorption to Aerosal Particles. The least is known about this process. In special cases, such as oxidation of SO_2, this process may be very important; however for most organic compounds, this process is believed to be relatively unimportant, although nonvolatile organics maybe emitted sorbed to the particulate in combustion processes.

Thus the physical processes are believed to be only of minor importance in determining the fate of organic compounds, and the three chemical processes discussed are expected to be the dominant processes in most cases.

SAR PROCEDURES

SAR methods have been recommended for the three dominant reaction processes listed above. First for the reaction of OH radical, the procedure is based on the group additivity principle (Benson, 1976). The method identifies the reactivity of various primary groups, such as primary, and tertiary carbon hydrogen bonds, and the effects that various substitutents have on these reactivities. The rate constant (kYOH) is evaluated by the expression:

$$k_{OH} = \sum n_i \alpha_i \beta_i k_i + \sum n_j \alpha_j k_j + \sum n_l \alpha_l k_l$$

The terms on the left in order represent hydrogen-atom abstraction, addition to olefins, and addition to aromatic rings. The n values are the number of times a given group in the molecule: α and β values are the factors by which substituent on the central (α) and adjacent (β) position affect the reactivity. The k values are the reactivity of each group with only hydrogen substituents. The method predicts rate constants within an accuracy of a factor of ±3. For more details as well as the data to apply to this equation see Hendry (1979, 1980).

The estimation of ozone reaction rate constants is limited to reactions with olefins and highly substituted aromatic compounds. One estimates the k values by matching the structure of the pollutant with similar structures of compounds with known reactivity. Care must be taken in determining reactivity of compounds with nonhydrogen substitution. For most compounds this method is expected to give rate constants with an accuracy of a factor of ±3. (Hendry 1979, 1980).

The proposed method for analyzing photolysis rate is to calculate the rate of absorption of light on a clear day, usually the summer solstice. Then a maximum possible rate of reaction is estimated by assuming a quantum yield for chemical reaction of unity. In most cases this maximum limit is much slower than the rate with OH or O_3. Thus it is possible to eliminate the possibility of photolysis being a chemical process. In cases where the maximum limit rate is comparable or greater than the sum of OH and O_3 reaction, then it is not possible to say how important photolysis will be and an experimental determination of the quantum yield will be necessary.

Research Needs

The prediction methods for OH and OH3 reactivity should be made applicable to a wider range of compounds. This will involve studying the reactions of both OH and O_3 with compounds with a wide variety of chemical substituents. Some of the substituents for which more data is needed are Cl, F, CN and NO_2.

REFERENCES

Hendry, D.G., 1980 Atmospheric Chemistry in T.M. II, et. al., Laboratory Protocols for Evaluating the FAte of Organic Chemicals in Air and Water, EPA-68-03-2227

Hendry, D.G., and R.A. Kenley, 1979, Atmospheric Reaction Products of Organic Compounds, Final Report, EPA-560/12-79-0001.

Benson, S.W., 1976, Thermochemical Kinetics, John Wiley and Sons, N.Y.

ENVIRONMENTAL FATE PROCESS 7: TRANSPORT OF CHEMICAL SUBSTANCES IN THE ENVIRONMENT by D. Mackay

Background

For assessing the environmental fate of chemical substances several classes of chemical transport property data are required in addition to the usual information about the magnitude of the sources, the receiving environment and prevailing environmental conditions which include reactivities, temperature and wind and water flow.

These classes are:

(1) Equilibrium Partition Data Between Media

These data can take several forms, for example, partition coefficients, Henry's Law constants or fugacity capacities. These data are fundamentally thermodynamic and contain no kinetic or time varibles. Although media may not be at equilibrium, this information is still needed because the rate of mass transfer between media such as air and water is usually dependent on the magnitude of the departure from equilibrium. Examples of such data are aqueous solubility, vapor pressure, octanol-water partition coefficient, sorption and biosorption partition coefficients.

(2) Transport Rate Terms Between Media

These data are usually in the form of mass transfer coefficients with units of velocity and can be conceived of as the velocity with which materials moves from medium to medium through the particular interface. Examples are volatilization mass transfer coefficients, bio-uptake half lives and sediment release rates.

We examine below the state-of-the-art in these two areas and, in particular, how structure activity relationships can contribute to quantifying the parameters.

Partition Data

The basic phase equilibrium information required is:

(i) Aqueous Solubility

A very powerful SAR for transport properties has proved to be the correlation of aqueous solubility with total molecular surface area, and with melting point, the latter being required to correct for the compound's fugacity ratio. This is well developed, examples being the various papers by Yalkowsky and Valvani (1980). Other approaches are to use carbon number of molar volumes (Mackay and Shiu, 1977). This approach may expose erroneous experimental data, for example for PCBs (Mackay et. al., 1980).

The success of theis SAR approach has been limited to specific homologous series which present the water matrix with organic

2-17

molecules of similar character (e.g., aromatics or alkanes or halogens). It would be very desirable to develop techniques to calculate solubility for molecules containing groups of differing character. This is judged to be achievable, albeit with a high degree of empiricism. This confidence arises from the success of K_{ow} correlations and the knowledge that K_{ow} and solubility are intimately related, indeed, to the extent that it is now clear that one measurement may substitute for the other, or more desirably, both are measured and their consistency checked. The latest solubility K_{ow} relationship has been published recently by Mackay et. al., (1980).

(ii) Vapor Pressure

There is at present inadequate treatment of the relationship between structure and vapor pressure for compounds of environmental interest. The most significant problem is the lack of reliable vapor pressure data for high molecular weight compounds. Approaches used for industrial chemical processing purposes seem appropriate by which vapor pressure can be estimated from normal boiling point, critical volume, and temperature (Reid et. al., 1975).

(iii) Henry's Law Constant

The air-water partition coefficient of Henry's Law Constant (H) is usually regarded as the ratio of vapor pressure to aqueous solubiltiy. It can thus be regarded as a combination of these two more funadmental quantities discussed above, rather than as a fundamental quantity in itself. Correlations as SARs thus tend to be of these two quantities. The only noteworthy point is that H is remarkably constant for a homologous series of organics of varying carbon number. Increasing carbon number tends to decrese both solubility and vapor pressure by the same factor thus H remaining fairly constant (Mackay and Shiu, 1981).

(iv) Octanol-Water Partition Coefficient (K_{ow})

K_{ow} is widely and successfully used to characterise hydraphobility, the leading proponents being Hansch, Leo and co-workers as discussed elsewhere in this report. Little need to be added here except to note that considerable effort has been devoted to calculating K_{ow} from molecular structure. K_{ow} thus acts as a "stepping stone" from molecular structure to variables such as partition coefficients for sorption or biosorption.

(v) Sorption Partition Coefficient (K_p)

It has been convincingly demonstrated that the principal sorbing site in soils and sediment for hydrophobic organics is natural organic matter which is probably humic or ligneous in character. Further, if K_p is expressed on an organic carbon basis (K_{oc}), where K_{oc} is K_p/(fraction organic carbon), it becomes possible to relate K_{oc} directly to K_{ow}. This approach, developed by Karickhoff, et. al., (1979) is invaluable as a predictive method for values of K_p.

(vi) Bioconcentration Factor K_b

This quantity, which is the biotic analog to K_p has been well documented and correlated with K_{ow}. There are examples in the many reports from authors such as Neely, Kenega, Branson, Freed, Chiou and others discussed in detail elsewhere in this report.

Transport Coefficients

The conventional approach for calculating the rate of transfer between media such as air and water is to apply equations based on Fick's First law of Diffusion which postulates that the flux N mol/m^2s is the product of a molecular diffusivity $D(m^2/s)$ a concentration difference $\Delta C(mol/m^3)$ and divided by a diffusion path length Z (m)

i.e., $N = D\Delta C/Z$

In practice, the diffusing molecule may experience a diffusive resistance or regime or either side of the interface and it is thus necessary to add two series resistance terms. It is often preferable, especially in the presence of turbulence, to combine (D/Z) as a single parameter, the mass transfer coefficient (K). Correlations are available for K between various media, usually as a function of the local level of turbulence as characterised for example by windspeed or river velocity. The transfer process may involve turbulent transfer in eddies which are exposed to unsteady-state transfer. The net result is that the K values tend to be proportional to diffusivity to a power lower than unity and generally in the range 0.5 to 0.67.

Adequate correlations are available for diffusivities in air and water as a function of the molecular volume or weight (Reid et al., 1975). There is thus little potential for SARs to contribute further to transport rate prediction. The major problem is the characterization of the turbulence near interfaces.

Research Needs

It would be invaluable to measure, gather and critically review physical chemical data such as solubility, vapor pressure, octanol-water partition coefficient and Henry's Law Constant for selected hydrophobic organic compounds covering a wide range of molar volume. These data would be useful for "calibrating" new methods of property mesurement such as HPLC techniques for K_{ow} measuresment.

At present there is a lack of reliable physical chemical property data for hydrophobic organics with which to test proposed SARs.

An achievable goal is that of developing SAR to predict these quantities from a knowledge of molecular structure alone. It is difficult to predict the ultimate accuracy attainable but considerable improvements are clearly possible.

REFERENCES

Mackay, D. and Shiu, W.Y., J. of Chem. and Eng. Data, 22, 4, 399-402, 1977.

Mackay, D., Bobra, A., Shiu, W.Y., Chemosphere 9, 701, 1980.

Mackay, D., Mascarenhas, R., Shiu, W.Y., Valvani, S.C., and Yalkowsky, S.H., Chemosphere, 9, 257-264, Pergamon Press Ltd., 1980.

Reid, R.D., Prausnitz, J.M., and Sherwood, T.K., "The Properties of Gases and Liquids", McGraw Hill, New York, 1977.

Mackay, D. and Shiu, W.Y., "A Critical Review of Henry's Law Constants for Solutes of Environmental Interest", J. Phys. Chem. Ref. Data (in press).

Karickhoff, S.W., Brown, D.S. and Scott, T.A., Water Res., 13, 241 (1979).

GROUP IV

TABLE OF CONTENTS

Page

4.0 PREDICTING ENVIRONMENTAL EFFECTS 1

4.1 Executive Summary 1

4.2 Introduction 9

4.3 State of the Art 10

  4.3.1 Present Concepts in Quantitative Structure Activity Relationships (QSAR) 10

  4.3.2 Availability of Appropriate Environmental Effects Data 12

  4.3.3 Estimating Effects Data by Interpolation 14

  4.3.4 Estimating Bioconcentration in Aquatic Organisms 18

4.4 Recommended Five Phase Research Program for Ecotoxicological QSAR 20

  4.4.1 Phase I--Selection of Representative Chemical Test Sets 20

  4.4.2 Phase II--Generation of QSAR Data Base: Ecologically Meaningful Endpoints 24

  4.4.3 Phase III--Consolidation of Test Sets into Rational QSAR Groups 27

  4.4.4 Phase IV--Refinements and Improvements 28

    4.4.4.1 Confidence Limits on Predicted Values 28

    4.4.4.2 Improvement in the Derivation of ECi 29

    4.4.4.3 Validation Experiments: Representative Benchmarks for Each Rational QSAR Group 30

4.4.5 Phase V—The Role of QSAR in Hazard Assessment: Extrapolations of Endpoints Values to Ecological No-Effect Concentrations 31

4.4.5.1 Species to Species Extrapolations 32
4.4.5.2 Acute Lethality to Chronic No-Effect Concentrations 33
4.4.5.3 Laboratory Data to Field Observations: 20/20 Hindsight 35
4.4.5.4 Environmental Decisions 36

2-2-5

LIST OF FIGURES

FIGURE NUMBER | Page

1 The Role of QSAR in Hazard Assessment 8

4.0 PREDICTING ENVIRONMENTAL EFFECTS

4.1 Executive Summary

There was general agreement that ecologically meaningful relationships between chemical structure and biological activity can and should be developed as tools for the prioritization of chemicals according to their degree of hazard in the environment. Although quantitative structure-activity relationships (QSAR) were viewed by this working group as inadequate in present form for regulatory decisions except where scientifically based interpolations are possible, the working group agreed that development of QSAR:

1. Provides a focused research target on a national problem and can be addressed effectively by many aquatic toxicological laboratories (government, industry, and academia) generating a minimum data base on a small but representative number of chemicals (rational groups)
2. Is a cost-effective approach to help identify the spectrum of most-to-least hazardous chemicals but only after the necessary QSARs have been built and validated.

Until meaningful QSAR relationships are developed and validated, the historical process of prioritizing chemicals used by various academic institutions, industrial scientists and regulatory agencies should remain in practice. Specifically, these groups assign chemicals to a status of "low priority" based on various combinations of the following exclusion principles:

1. If the potential for exposure is low based on the environmental release rate
2. If the chemical is unstable in water based on chemical predictions,
3. If the chemical has been looked for and not found in the environment
4. If the soil-water-air equilibrium causes minimal concern, e.g., fugacity calculations based on structure
5. If the water solubility or molecular weight preclude significant accumulation, or
6. If in the wisdom of a panel of interdisciplinary experts the chemical is judged to be of low concern to man or the environment

Since the objective of environmental QSAR development is to prioritize chemicals based on hazards, defined as the margin of safety between a no-effect concentration and an exposure concentration, this Working Group recommended no research and development or environmental decisions based on qualitative structure activity relationships, i.e., those based on pass-fail test results, or those with a questionable dose-response, or those with a nonquantitative dose-response, or those where the endpoints are not well defined, characterized, or are of unknown ecological significance.

The working group recommended the following Five Phase Research Program be initiated as soon as possible:

PHASE I--Select about 1200 Chemicals, Divided into 60 Tests Sets, Representing All Relevant Chemicals.

The 1200 chemicals should be selected from a large list of chemicals including industrial chemicals, naturally occurring chemicals, drugs and pesticides which remain after a first-pass exclusion by assignment to "low priority" status. Also the 1200 chemicals would include those "research" chemicals needed to complete a specific test set. Each test set should contain at least 20 chemicals which span a wide range of values described by molecular descriptors. The 1200 chemicals were chemicals/test sets x 60 test sets.

The selection of the 20 chemicals in each test set should be based on (A) chemical structure similarity, (B) at least 10 chemicals with "well behaved functional groups" (characterized by weak hydrogen bonding) and 10 chemicals with functional groups of unknown chemical behavior and (C) chemicals with a similar mode of action according to the collective judgments and assumptions of medicinal chemists. For example, methanol and ethanol are similar chemically but might belong to different test groups because their mode of action is different.

PHASE II--Generation of QSAR Database: Four Ecotoxicologically Meaningful Endpoints.

A. Lethality to aquatic vertebrates measured with fathead minnows in the 96 hour LC50 test.

B. Immobilization effects to aquatic invertebrates measured with Daphnia magna in the 48-hour EC50 test.

C. Chronicity factor (a modification of the conventional application factor), determined for several chemicals in each test set.

D. Lethality to mammals measured with rats in the seven-day LD50 test (diet or water).

These endpoints were selected for the following reasons:

1. Ecological importance--predictive relationships between these endpoints and other ecotoxicological data can be established. These other data include predicted lethal concentrations to a second aquatic species based on fathead minnow data, a predicted chronic no-effect concentration based on a chronicity factor and acute data, predicted field observation or aquatic epidmiological data based on laboratory data, and predicted no-effect concentrations to terrestrial animals based on the rat data.

2. Mammalian toxicology In addition, to the need to identify chemicals that cause harm to terrestrial animals, mammalian

toxicological QSAR were also included in the recommendation to emphasize the need and the mutual benefits of cross-fertilization between mammalian and environmental toxicology.

3. *Standard test methods* The test procedures needed to generate this QSAR data base are reasonably well developed, but it was the intent of this working group for specific protocols to be drafted for each of these special QSAR tests to assure the highest quality data possible and to define the interlaboratory accuracy and precision. These protocols would specify several measurements which are not routinely made, including gross pathological observations for the most sensitive toxicological response, slope of the dose response and time to effect data. Also, protocols for water solubilities and octanol: water partition coefficients (at *infinite dilution*) were recommended since correlations with these chemical properties are anticipated. These maximum effort protocols should *not* be considered necessary for testing specific chemicals for conventional hazard assessments in the future, but are necessary for building these QSARs for ecological extrapolations.

PHASE III--Consolidation of the 60 Test Sets of Chemicals into a Minimum Number of Rational Groups.

The process of lumping and splitting chemicals into ecologically meaningful rational groups was seen by this work group as the largest challenge before using QSAR for prioritizing chemicals. This challenge should be greatly aided by computerized data analysis of the various combinations and permeations of molecular descriptors and toxicological endpoints to identify the most reliable regressions.

PHASE IV--Refinements and Improvements

New molecular descriptors, new endpoints, confidence limits of predictions, accuracy criteria for the use of QSAR in environmental decisions, and validation of at least one benchmark chemical for each rational group of chemicals were areas needing refinement and improvement after the rational group were identified. Working groups II and IV both agreed that developing new molecular descriptors was best accomplished as a feedback process after the endpoints data had been generated. Many possible new endpoints could be evaluated including predictors of the rate of metabolic toxification and detoxification in aquatic organisms, time-independent EC1 values (concentrations causing 1 per cent of some effect), and bioconcentration potential for chemicals not fat soluble

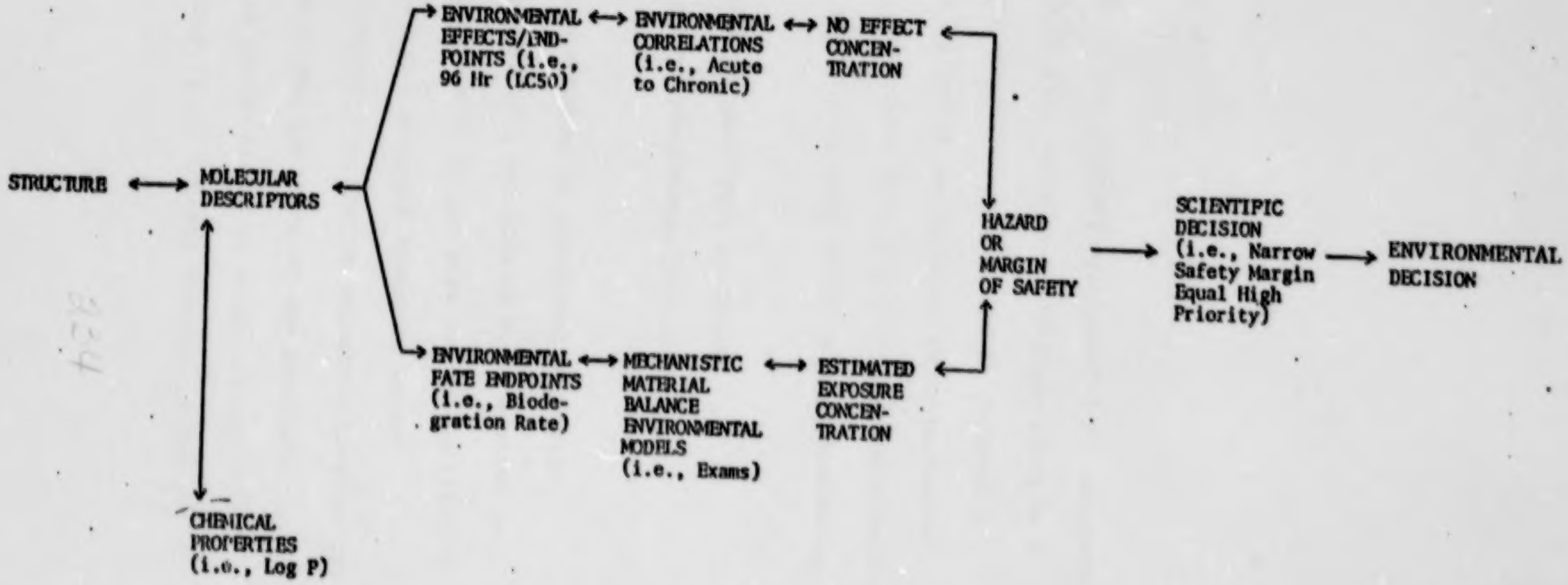
The uncertainty associated with each of the various predicted values will get larger with each extrapolation. Therefore, it is important to assign confidence limits to each predicted value and to prepare an accuracy criteria for accepting or rejecting individual values. Comparisons of the values for untested new chemicals with those values of benchmark chemicals of similar structure improves the confidence of the scientist for a specific predicted value. Furthermore, if a new chemical fails to fit into one of the rational groups for using QSAR, the scientist can draw on other sources for wisdom to help place the new chemical in its proper priority.

PHASE V--Role of QSAR in Hazard Assessment: Extrapolations of Endpoint Values to Ecological No-Effect Concentrations

Figure I summarizes the role of QSAR in the overall process of prioritizing chemicals. The process starts with a chemical structure and ends with an estimate of hazard and possibly an environmental decision. The key to this viewpoint was the need to estimate a toxicological endpoint value from regression equations based on a calculatable descriptions of the chemical structure, i.e., molecular descriptors.

Kee

FIGURE 1. THE ROLE OF QSAR IN HAZARD ASSESSMENT


STRUCTURE
MOLECULAR DESCRIPTORS
CHEMICAL PROPERTIES (i.e., Log P)
ENVIRONMENTAL EFFECTS/END-POINTS (i.e., 96 Hr (LC50)
ENVIRONMENTAL CORRELATIONS (i.e., Acute to Chronic)
NO EFFECT CONCEN-TRATION
ENVIRONMENTAL FATE ENDPOINTS (i.e., Biode-gration Rate)
MECHANISTIC MATERIAL BALANCE ENVIRONMENTAL MODELS (i.e., Exams)
ESTIMATED EXPOSURE CONCEN-TRATION
HAZARD OR MARGIN OF SAFETY
SCIENTIFIC DECISION (i.e., Narrow Safety Margin Equal High Priority)
ENVIRONMENTAL DECISION

4.2 Introduction

A common need of the regulator and regulated is to focus testing resources on those chemicals with the greatest potential for causing harm to man and the environment. Historically, the process of prioritizing chemicals for testing has included consideration of toxicological potency, estimated exposure, persistence, and bioaccumulation potential. The cost of generating information needed for even the simplest evaluations of the large numbers of chemicals is a national problem and has become the driving force for exploring relationships between chemical structure and those properties or effects considered essential for environmental hazard assessment.

Specifically, this working group addressed the relationships between structure or physical properties and the toxicity to aquatic organisms. Because of its importance in toxic effects, bioaccumulation potential in aquatic organisms was also considered by this Working Group. The working group recognized the need for parallel reviews of relationships with mammalian toxicological effects including carcinogenicity, mutagenicity, and teratogenicity. It was also felt by the group that too little is known about the uptake and transformation of chemicals by plants to consider the applicability of QSARs at this time. It is intended that the relationships discussed in this section be used only as a part of an overall hazard assessment.

4.3 State of the Art

4.3.1 Present Concepts in Quantitative Structure Activity Relationships

A large body of quantitative structure activity relationships (QSAR) analyses has evolved from studies in medicinal chemistry over the last two decades. The principle purpose of these studies was the attempt to elaboratemolecular mechanisms of drug action and to predict new leads in drug design. Out of these investigations has come a paradigm of methodology based upon the awareness that molecular structure generates molecular function or biological activity.

These studies have revealed the value of describing molecular structure in a form suitable for correlational analyses with biological data. Molecular structure can be mirrored basically in two ways, leading to molecular descriptors useful for SAR analyses.

The first class of molecular descriptors can be called the quantitation of molecular structure. This classification includes the most fundamental description, quantum mechanics, or on the more approximate form, molecular orbital theory. This provides analytical expressions of electron probability and energy which can be translated into a description of models predicting dynamic events in which the molecule may take part. Numerous examples appear in the literature and a book covers the more elementary aspects (Kier, 1971).

At an intermediate level of rigor, a molecular structure may be quantitated by using a method known as molecular connectivity (Kier and Hall, 1976). This method results in indexes derived from the structure encoding information about branching, unsaturation, cyclization, and heteroatom content. In addition to these more widely applied methods to biological SAR, other purely structural descriptors have been used based upon calculated volume and symmetry.

A second broad class of methods used in SAR analyses has employed physical properties as descriptors of molecules, for use in correlations with biological activity. Chief among these has been the use of the partition coefficient, Log P, introduced and developed by Hansch and Colleagues (see Group I discussion). A large amount of evidence has accumulated, leading to the belief that this physical property mirrors limiting steps in many biological phenomena, permitting its use in correlational analyses and predictions.

Other physical chemical properties of value as molecular descriptors for SAR analyses have been developed by Hammett and Taft, including the molar refraction, ionization constants, and reactivity ratios formulated into electronic and steric substituent constants (see Group I discussion).The development, general availability and record of success of these SAR methods recommends them as initial approaches to the SAR analyses of the relative toxicity of molecules in the biosphere.

4.3.2 Availability of Appropriate Environmental Effects Data

These data are discussed in context of chemicals, species of aquatic fauna, type of toxicity data, and possibilities of rapid extensions of the data base.

Chemicals. A very extensive data base is available for pesticides. The data base is being expanded automatically, since registration applications must contain aquatic toxicity data. Since pesticides tend to act by specific mechanisms, in contrast to the generally nonspecific action of general industrial organic chemicals, SAR's for pesticides should be considered separately. A quite extensive data base is also available for surfactants, although much of the data is scattered through the literature.

The situation is much worse for general organic chemicals. Extensive data are available only for certain classes of compounds, such as phenols and onlyi a few papers deal with more than several chemicals (see for example Crisp et al. 1967; Price et al. 1974; Dawson et al. 1975/77; Schultz et al. 1980; McLeese et al. 1979; and Bridie et al. 1979).

A considerable effort is needed to retrieve continuously data on general organic chemicals from the literature. ERLDuluth is attempting to develop SAR's within a set of 15,000 chemicals. There seems to have been little progress on the computerized data base AQUATOX, carried as a part of CIS.

The majority of the toxicity data is expressed in terms of LC50, usually for a 96h exposure, but other exposure times are encountered as well, most frequently 24 and 48h. Some data are in terms of lethal thresholds (or incipient lethal concentrations).

The data usually give only a single point (50% mortality) on a response curve and the varying exposure times complicate the situation even further.

While the predicting of lethality (toxicity) has been related to a number of structural and physical parameters (i.e., number of carbon atoms, alkyl-chain length, alkyl-ring addition, molecular weight, boiling point, water solubility, as well as various substituent constants), the Log P, which is often colinear with most of the above mentioned parameters, is at present the most effective predictor especially with cogeneric series (Hansch and Gleve 1971; Koppermann et al. 1974; Zitko et al. 1976; Hutzinger et al. 1978; McLeese et al. 1979; Schultz et al. 1978/1980). The literature suggests that structure-toxicity relationships have been successfully developed for groups of chemicals divided along classical organic chemistry lines (i.e., aliphatic, alicyclic, aromatic, heterocyclic), with pesticides and surfactants having to be further subdivided and studied separately. General trends that are presently available suggests that aliphatics have relatively low toxicities with the smaller and more reactive compounds being most toxic. Aromatic compounds are generally more toxic than their aliphatic analog. The toxicity of the mono- and diaromatics is usually increased by halogen, nitro, and alkyl substitution. While fewer hetrocyclics have been studied, the data to date suggest that nitrogen substitution decreases the toxicity in reference to the corresponding aromatic hydrocarbon.

4.3.3 Estimating Effects Data by Interpolation: Summary of Structure-Aquatic Toxicity Relationships

The aquatic toxicity database is largely comprised of toxicity tests involving metals and pesticides. It is concluded that reliable data on fewer than 500 of the 45,000 organic industrial chemicals have been reported. Extensive data are available only for certain classes of compounds such as phenols, and only a few papers deal with more than several chemicals (see for example Crisp et al. 1967; Price et al. 1974; Dawson et al. 1975/77; Schultz et al. 1980; McLeese et al. 1979; and Bridie et al. 1979).

Although this is not intended to be an exhaustive review, a few studies can be cited to illustrate the potential of structure-toxicity relationships with aquatic organisms. Zitko et al. (1976) demonstrated that the toxicity of alkyl mitrophenols to Atlantic salmon can be estimated by the relationship:

$$\log \frac{1}{C} = 3.31 + 0.31 \text{ Log P}$$

McLeese et al. (1979) showed that the toxicity of phenols and anilines to shrimp could be estimated by:

$$\log \frac{1}{C} = 2.50 + 1.02 \text{ Log P}$$

Koppermann et al (1974) found that the toxicity of a wide variety of phenols to Daphnia could be estimated by:

$$\log \frac{1}{C} = 3.73 + 0.53 \quad + 0.53 \text{ F} + 0.64 \text{ R}$$

Hutzinger et al. (1980) found that a variety of chemicals were toxic to Atlantic salmon according to the relationship:

$$\log \frac{1}{C} = 3.13 - 0.26 \text{ Log P}$$

Veith et al. (unpublished) found that the fathead minnow LC_{50} for ethers, alcohols, ketones, alkyl halides, and chlorinated benzenes could be estimated by:

$$\log \frac{1}{C} - 1.17 + 0.94 \text{ Log P}$$

From this same toxicity data, Kier and Hall (unpublished) developed a structure-toxicity model using the connectivity indices as follows:

$$\text{Log}_{\text{LC50}} = 0.28 - 1_{x}^{v}$$

where $1_{x}v$ is the first order valence connectivity of the industrial chemicals. This latter approach has considerable potential since the connectivity indices can be quickly computed from structure.

Schultz, (unpublished) found the acute toxicity of 24 nitrogen heterocyclic chemicals to _Tetrahymena pyriformis_ to be estimated from the equation:

$$\text{Log}_{\text{LC50}} = 0.71 \text{ Log P} - 1.24$$

The data indicate that for structurally related groups of chemicals the toxicity can be predicted. The accuracy of the prediction depends, of course, on the quality of the data base and on the similarity of the compounds whose toxicity is predicted to the compounds in the data base. It is not possible as yet to estimate the toxicity of just any organic compound, and "across the board" toxicity relationships may never be derived, although a group of experts may be able to rank chemicals into certain toxicity classes. Pesticides and surfactants will have to be handled separately. Pesticides will have to be subdivided into classes such as organophosphates, carbamates, pyrethroids, etc. At the same time, the development of structure-toxicity relationships for pesticides may be of less significance because it is likely that actual toxicity data rather than predictions will be required for registration.

It appears that many aliphatic compounds have generally low toxicity (96-h LC50 more than 30 mg/L). The more toxic compounds are mostly small, quite reactive molecules.

Aromatic compounds, limited at this stage largely to substituted benzenes and napthalenes appear generally more toxic than aliphatic compounds. The toxicity is usually increased by halogens, nitro groups, and reactivity, and is decreased by sulfogroups.

Too few alicyclic and heterocyclic compounds have been studied to make conclusions about their toxicity. From the data, pyridine derivatives are certainly less toxic than the corresponding benzenes.

Among surfactants, acute fish toxicity and primary biodegration rate data have been found to correlate with experimentally measured octanol/water partition coefficients and with the basic chemical structure (alkyl chain length) of the molecules.

Homologous series of surfactants demonstrate increased acute fish toxicity as the length of the alkyl chain is increased. This increase in toxicity is five- to ten-fold for each additional two alkyl carbons. Similarly, the octanol/water partition coefficient and solubility in water can be directly correlated to toxicity within the homologous series. A maximum in acute fish toxicity was observed at an intermediate alkyl chain length and was also correlated with partition coefficient.

Data developed for a wide variety of homologous series of surfactants taken from anionic, cationic, nonionic and Zwitterionic surfactant classes indicate that significant correlations exist between acute fish toxicity and the hydrophobic nature of the molecules. Regression equations developed between acute fish toxicity and water solubility or partition coefficient can be used to estimate the acute toxicity of related surfactant structures thus demonstrating the predictive utility of SARs for safety evaluation of new and existing surfactants.

It is tempting to explore the possibility of extending the database. For example, by establishing correlations between aquatic and mammalian toxicity data and thus exploiting the extensive mammalian toxicity database. This approach has been suggested for example by Seba (1975). It appears feasible for some classes of structurally related compounds. The correlations are likely to improve by using mammalian inhalation toxicity data since these are closest to the aquatic toxicity data.

4.3.4 Estimating Bioconcentration Potential in Aquatic Organisms

The extent to which organic chemicals bioaccumulate in aquatic organisms is dependent on many membrane transport and metabolic processes. Since many of these processes covary with hydrophobicity of chemicals, the bioaccumulation potential can be estimated by linear equations involving Log P. For example, Veith et al. (1979) showed that the bioconcentration factor for chemicals with fathead minnows could be estimated by:

$$\text{Log BCF} = 0.85 \text{ Log P} - 0.70$$

Similar relationships were reported earlier with more limited data sets by Chiou et al. (1977) and Weely et al. (1974).

In the present state of the art, bioconcentration of organic chemicals in aquatic fauna is more predictable than lethality. The bioconcentration factor (concentration in biota/concentration in water, increases with increasing octanol-water partition coefficient Kow (or with decresing solubility in water) and a number of such

2-44

relationships have been published. Chemicals with Log Kow >4 have a potential to be bioconcentrated. The published relationships tend to overestimate bioconcentration of chemicals with very high calculated values of Kow (Log Kow >6). This discrepancy may be caused by erroneously high calculated values of Kow, by nonlinearity of the relationship between Log (BCF) and Log Kow, or by extremely low solubility in water. It appears that highly lipophilic crystalline organic chemicals with MP >300°C are not taken up by fish from water. This problem deserves a further study.

The uptake and depuration rate constants of the frequently used one-compartment model are also predictable from octanol-water partition coefficients.

Adsorption coefficients on suspended solids and bottom sediments are predictable from octanol-watef partition coefficients and, consequently, the bioconcentration of chemicals in the presences of suspended solids or sediments is predictable.

The relative importance of various routes of uptake may be judged by considering the terms of the equation

$$\frac{dC_b}{dt} = K_w C_w + K_f C_f - K_e C_b$$

Where K_w, K_f, K_e = Rate constants for uptake from water, food, and for exretion, respectively C_b, C_w, C_f = Concentration in biota, water, and food respectively.

4.4 Recommended Five-Phase Research Program for Ecotoxicological QSAR

As indicated by the review of the state of the art, toxicity and other environmental effects of chemicals are predictable when a sufficient data base is available. It was the consensus of the working group that a data base representing a wide range of industrial organic chemicals should be developed. The estimated number of chemicals in this data base is 1200, and the project will consist of five consecutive phases described below.

4.4.1 PHASE I--Selection of 1200 Chemicals for QSAR for Ecological Toxicity

The number of chemicals which might become candidate for structure activity analysis for ecological toxicity is extremely large, and it therefore needs to be examined in a structured approach.

Certain chemicals should be excluded from consideration for SAR analysis for aquatic ecosystems. Substances which are readily and rapidly hydroloysed and substances which have a very high vapor pressures are candidate for exclusion from SAR analysis for aquatic ecosystems since they are unlikely to persist long enough in this ecosystem to produce effects under most realistic conditions.

The remaining substances should separated initially into five major groups: aliphatic compounds, alicyclic compounds, heterocyclic compounds, aromatic compounds, and surface-active compounds, each of these groups being divided further into several subgroups.

About 20 conditions per subgroup should be used to develop SAR's. After the development of SAR's, it is likely that some subgroups may be combined or that additional subgroups may have to be formed, depending on the predictability of effects by SAR's.

In parallel of this process it may be possible to classify chemicals by thetype of biological action mechanism.

Biological effects produced by all chemicals can be roughly classified into three major cause-effect relationships. Firstly, the effect may be nonspecific action on biological membranes, resulting in behavioral effects, hyperexitability, narcosis, and eventually death. Effects on membrane bound enzymes and election transport mechanisms may also be produced. Effects such as these are most likely correlated with the octanol/water partition coefficient.

Secondly, effects may be nonspecific due to a direct reaction of the chemical with functional groups of proteins such as R-NH_2, R-COOH, or R-SH. These effects would be expected to produce pathological effects at the port of entry, and the major sites of metablism, and at the site ofexecution. Such effects would most likely correlate with indices of molecular reactivity, or direct descriptors of molecular structure through the partition coefficient might still play an important role, since it is related to the ease with which the chemical may cross cellular membranes.

Thirdly, effects may be undue to specific interactions with receptor sites and reactive centers of enzymes. Currently, specific interactions are less predictable from QSAR than nonspecific interactions. Specific interactions can further be subclassified according to Hancsh's opening remarks at this workshop, into chemicals blocking information generation and transmission and chemicals blocking energy generation and consumption. Most frequently, chemicals which produce such effects are analogues of the naturally occurring biochemicals. Thus, ecogenous chemicals which resemble the endogenous biochemicals may exert unusual effects an organisms by producing enzyme inhibition, participating in lethal synthesis, or interactions directly iwth physiological receptor sites. The toxicological results can be pronounced and with the present state of the art, are often unpredictable. Furthermore, the relationship of a xenobiotic to a receptor site or reactive center on an enzyme is not often obvious from structure.

It is also important to identify structural analogues of biochemicals since these may have toxic effects not predictable by SARs, such as blocking crucial enzymes or metabolic pathways. Various ethanolamine derivatives are an example.

In addition, it may be fruitful to examine effects of different substituent groups on the SAR within homologous series to determine whether the variant substituent group produced an unusual effect on the biological activity of the compound. This approach could eventually lead to a rational SAR analysis of compounds which do not readily fit into a homologous series.

It is very likely that a chemical will stand out, in the judgment of a medicinal chemist, as being very potent in terms of potential mode of action. This judgement or medicinal chemical wisdom would be based upon broad experience with different classes of molecules as 1) medicinal agents used in chemotherapy, 2) research chemicals identified as being potent but not otherwise useful in therapy or 3) a general knowledge of the adverse of prominent side effects of a molecule. This professional judgement may permit the categorization of certain molecules into a class based upon a judgement of a common mechanism of eliciting toxicity. Thus in this category may be found molecules of disparate structures but with a presumed common receptor mechanism producing toxicity. The corollary is true that similar chemical structures may be classified into different categories based upon mechanism of action at different receptors.

It is also important to flag structural analogues of biochemicals since these may have toxic effects not predictable by SAR's, such as blocking crucial enzymes or metabolic pathways. Various ethanolatine derivatives are an example.

In addition, it may be fruitful to examine effects of different substituent groups on the SAR within homologous series to determine whether the variant substituent group produced an unusual effect on the biological activity of the compound. This approach could eventually lead to a rational SAR analysis of compounds which do not readily fit into a homologous series.

4.4.2 Phase II--Generation of QSAR Database: Ecologically Meaningful Endpoints

One of the first rules of structure-activity research is that meaningful and well-defined endpoints for biological activity be identified. Although there are many biological responses which are used to characterize the toxicity of chemicals, the work group concluded that the fathead minnow (Pimephales promelas) 96-hour LC50 is the best characterized measure of acute lethality to aquatic vertebrates, the 24- and the 48-hour EC50 (immobilization) to Daphnia magna is the best characterized effect on aquatic invertabrates, a modified application factor (chronicity factor; same effect acute and chronic exposure), is the best estimator of chronic toxicity to aquatic organisms, and the rat (Rattus norvegicus) 7-day LC50 with the toxicant administered in water or food is the best estimator of acute lethality to mammals.

It was noted that in addition to the above "single point" responses, other responses or observations obtained during the tests should be recorded and used, if possible, in additional SARs. For example, in lethality tests with fish, the lethality curve (LT_{50} vs. concentration) may be characterized by curve-fitting and the parameters used in SARs.

At high concentrations, if administered for sufficient duration, all of the test chemicals are expected to produce deaths. This effect is nonspecific, does not by itself produce any indication of mechanism of toxicity, but is largely measured and statistically quantified. Thus, mortality presents a simple means of comparing the same end-point of toxicity for all SAR analyses, without regard to mode of toxic action. Thus, in spite of its recognized shortcomings, it was selected as a major end-point for acute toxicity to be studied in 96-hour LC50 or EC50 protocols in fathead minnows, and 7-day dietary or drinking water LC50 protocols in rats. The species were selected on the basis that they provided the absolute minimum sample of animal species which had predictive value for both aquatic and terrestrial ecosystems. In the aggregate, plants and microorganisms tended to be less sensitive than the species selected. The 7-day LC50 protocol in rats was selected in favor of the routine oral LD50 in rats because the administration of chemicals by gavage in the latter test was felt to produce such unusual pharmacokinetic phenomena as to severly limit the predictive value of the LD50 test.

The committee was well aware of the limited information contained in choosing mortality as an endpoint. Consequently, other endpoints should be considered for routine examination, at least during the research phase of establishing SAR evaluations. Sensitive indicators in Daphnia, fish, and rats would be behavioral effects classified according to the narcosis-hyperexcitability spectrum. Additionally, fish should be examined for changes in body weight, lesions in gills, liver, and kidney, and whenever possible EC50 concentrations should be calculated for any effects, and these effects should be examined and compared to SAR values whenerver possible. Rats should also be examined for changes in body weight, lesions in the liver, kidney, and gastrointestinal tract. In addition, any unusual effects in organ systems or possibly in clinical chemistry, should also be noted in fish and rats.

While the initial research approach for the use of SAR analyses would be expected to explore the correlations between various molecular properties and descriptors with a limited number of endpoints and within many homologous series of chemicals, it should be considered that the response of the biological entities may follow a different ordering of SAR than that of homologous series within major groups. Thus, the toxic responses of the organisms tested, especially the unusual responses which have been noted during the test, should be tested against various arrays of molecular properties or descriptors to determine whether the biological organisms respond to chemical insult according to discernible patterns which differ from the ordered perspective of structural series.

4.4.3 Phase III--Consolidation of Test Sets in Rational QSAR Groups

Since the utimate objective of the project is the creation of the ability to predict a potential hazard status for an untested molecule, the ultimate goal of the QSAR effort must be directed toward the definition of such a molecule in quantitative structural terms. Thus a rational group with ecological meaning becomes a group of molecules with a markedly common functional group but varying in other sturctural features which mitigate the experimental or predicted toxicity. Thus a molecule predicted from a QSAR analysis as being toxic, must be expressed in quantitative structural terms which the chemial scientist can recognize. The expression of a molecular structure in terms of physical property creates uncertainity and ambiguity since we are not accustomed to defining a series of ketones in terms of increasing boiling point but in terms of an increasing count of the number of carbon and hydrogen atoms.

The use of statistical and mathematical procedures will permit the optimization of the structural information content of the QSAR analyses. These could take the form of discriminant analyses or pattern recognition beyond simple regression analyses. The ultimate goal is, of course, the definition of a structure, recognized by any scientist as depicting a molecule which, though untested, is predicted to possess a certain range of toxicity.

4.4.4 Phase IV--Refinements and Improvements

The accuracy of QSAR predictions of toxicity depends on the quality of the data base and on the similarity of the investigated chemical to those in the data base. It is likely that the accuracy of predictions will improve with improvements in the data base and with experience. Obviously, some of the predictions will have to be validated. For environmental decisions, the QSAR in predictions of toxicity must be supplemented by QSAR predictions of environmental fate, yielding expected environmental concentrations and their kinetic profile.

4.4.4.1 Confidence Limits on Predicted Values. Experience to date with aquatic toxicity data indicates that the precision of LC50 values is, under optimal conditions with a single test species, within a factor of 2(±100%). The accuracy of the LC50) values isaffected mainly by properties of the tested compound (solubility in water, volatility), type of test (flow-through, static, static with replacement), and the use of nominal or actual (measured) concentration. For many compounds the accuracy was about an order of magnitude, mainly because of poor testing protocols. With improvement in testing protocols, particularly by using actual rather than nominal concentrations, the accuracy is likely to improve the error within a factor of 2(±100%). The working group also reviewed data which suggests that the chronic toxicity of many chemicals to fish is greater than the LC50 because the endpoints measure different effects caused by the same general mode-of-action that results in death in the acute test rather than completely different mode-of-action at the chronic exposure level. An example is that lethality caused by narcosis will require 3 to 5 times the concentration needed to make larval fish lethargic enough to stop feeding normally. Consequently, the enbryo-larval test would be expected to produce a chronic no-effect endpoint 3 to 5 times lower than the LC50. The application factor is expected to be larger for metals and highly bioaccumulative chemicals than for water soluble organic chemicals where the application factor is quite small. Even moderately toxic chemicals such as pentachlorophenol have application factors of approximately two.

The working group recommends model development for relationships of acute and chronic toxicity so that the SAR data base and the predictive technology developed from it can be used to the fullest extent in assessing effects of chemicals.

4.4.4.2 Improvement in the Derivation of ECi. In the past, estimations of ECi (concentration causing i% of some effect) from acute data were done at one point in time, usually on the data observed on the final day of exposure. This approach is probably adequate for predicting chronic effects from water soluble chemicals since acute and chronic toxicity values are generally not very different. However, discrepancies between acute and chronic effects become large with highly cumulative chemicals.

One possible way of improving ECi data in predicting chronic effects is to estimate ECi values for three or more time periods within the acute test and to derive the infinite ECi by solving for "a" in the equation $\log ECi = a + b\ (^{1}/t)$ where "t" equals time. The dose-response curve for mortality should be a continuum in dose and time if the mode of action is the same under both acute and chronic test conditions. Also, a high partition coefficient may make little difference if adequate acute data are available in time.

4.4.4.3 Validation Experiments: Representative Benchmarks for Each Group. It seems that compounds whose molecular descriptors predict high toxicity within QSAR groups should be tested to validate the predictions. Additionally, compounds whose molecular descriptors hint at possible differences from the data base should be tested. The validation process may be conveniently incorporated into the preparation of the data base since the validation results become a part of the data base.

Problems may arise with relatively nontoxic chemicals and an upper limit for "realistic" concentrations should be established.

For a start, this limit may be set at 100 mg/l. Compounds yielding endpoints only above this concentration should not be used in data bases.

4.4.5 Phase V--The Role of QSAR in Hazard Assessment: Extrapolation of Endpoint Values to Ecological No-Effect Concentration

The consensus of the committee was that a minimum biological data base for a given chemical should include environmental toxicity information on an aquatic invertebrate, an aquatic vertebrate and a mammal. Consideration was also given to vascular plants and algae, but, as a first approximation, seemed reasonable to assume that environmental concentrations of chemicals that would in almost all cases protect aquatic vertebrates and invertebrates, would also protect aquatic plant life.

The recommendation to limit the number of test organisms to these three listed is parallel to current trends within U.S. and International (O.E.C.D.) testing programs for safety evaluation of new and expanded-use chemicals. As our background and data base for comparative toxicology with aquatic species have developed it has become increasingly evident that protection criteria based on laboratory toxicity data for fish and daphnia can be used to protect other species from marine and freshwater environments. For the purposes of screening, ranking or prioritizing chemicals with order of magnitude accuracy, "a fish is a fish is a fish." Until such time as the state of the art for aquatic toxicology develops to identify a more meaningful "white rat" species, these test species will remain as the most important species for decision making. On the other hand, for chemicals with narrow safety margins, toxicity to additional freshwater and marine species might be advisable. the kind of aquatic toxicity information the committee felt was needed includes: acute data, partial chronic data, and laboratory-to-field data. The following are the rationale for the test organism selection and elaboration on the tests methods:

4.4.5.1 Species to Species Extrapolations. The working group reviewed some evidence that suggests that lethality data derived from one aquatic species can be used to estimate the lethality to other species. Although a comprehensive comparative toxicology program is needed to define these relationships, the literature suggests that the relative sensitivity of different aquatic species to many organic industrial chemicals may follow consistent trends.

Kenaga (1978) that the LC50 for bluegills could be estimated from the LC50 of fathead minnows with the expression:

$$\text{Log BG LC50} = 0.01 + 0.82 \text{ Log FM LC50}$$

Also, LC50 to rainbow trout can be used to estimate the LC50 for fathead minnows with the expression:

$$\text{Log FM LC50} = 0.40 + 0.91 \text{ Log RT LC50}$$

and the rainbow trout LC50 is related to the bluegill LC50 by the expression:

$$\text{Log RT LC50} = -0.03 + 1.01 \text{ Log BG LC50}$$

Although data are insufficient to validate similar relationships with other species, it is commonly observed that channel catfish and Daphnia are consistently more sensitive than fathead minnows. It was also apparent that the range of sensitivities among aquatic organism may be small (perhaps less than a factor of six) for organic industrial chemicals compared to the range of toxicities even within a homologous series. Therefore, further development of the comparative toxicology of important species would greatly increase the value of SAR data bases.

4.4.5.2 Acute to Chronic Extrapolations During the process of safety evaluation of new and expanded-use chemicals, there is a need for the development of sequentially more refined estimates of toxicity to aquatic life until such a point where comparisons of known or projected biological effects can be effectively differentiated from the predictions of expected environmental concentration by acceptable margins of safety. This has created the need for early accurate estimates of chronic toxicity to aquatic life to guide manufacturing and regulatory decisions. Efforts to develop this early prediction of chronic effects have focused on the development of mathematical relationships between acute toxicity and chronic effects. This relationship is termed the application factor (AF) and is defined as:

$$\text{Application Factor} = \frac{\text{No Observed Effect Concentration}}{\text{96 hr. LC50}}$$

This relationship has been demonstrated to be a relatively efficient predictor (P = 0.76 to 0.99) of chronic toxicity when only 96-hour acute data are known, provided the AF has been developed for at least one species of fish. The AF is useful on a given compound between species of fish where LC50 species sensitivity varies greatly. AF's can be predicted by the use of equations for unrelated chemicals and the specified organisms, as well as for specified chemicals and unrelated fish species. AF's do not correlate well when compared between fish, daphnids and mammals.*

---

*E.E. Kenaga (1979) Aquatic test organisms and methods useful for assessment of chronic toxicity of chemicals (Fran: Analyzing the Hazard Evalation Process. K.L. Dickson, A.W. Maki and J. (Cairns, editors) Amer. Fish. So. p. 101-111.

4.4.5.3 Laboratory Data to Field Observations: 20/20 Hindsight. The ultimate purpose of conducting environmental effects studies on chemicals in the laboratory is to predict their effect in the real world. When the effect concentration of a chemical is known it is then assumed that adverse environmental impacts could be avoided. The primary questions that must be answered include the following: (1) are the current state-of-the-art laboratory tests being performed (i.e., the proper tests,) (2) are the responses to doses comparable in lab and field studies, and (3) do not structure/activity relationships and biological/physical/chemical properties of a chemical maintain their predictive usefulness from laboratory studies to the real world?

The experience we have on effects of chemicals in the real-world is limited to a very few chemicals. However, it was felt that acute and chronic toxicity (birds) and biological uptake data (plants, invertebrates, fish, mammals, birds) from the laboratory present a reasonable model of what could happen in real world ecosystems. Chronic laboratory studies were considered particularly useful for predicting such classical effects as mortality, growth, and reproduction. If other "nontraditional endpoints" were included in the laboratory chronic test, the group thought we might also have a tool to avoid some of the move subtle ecosystem function effects that adversely impact population distribution--abundance, primary and secondary production, and behavior. In other words, the responses for parameters which we can measure in the lab and in the field seem to yield similar conclusions (and if we could only "get smarter" we may have the tools to satisfy other important needs).

On the subject of whether the important properties of a chemical holdup from lab to field, the answer seems to be yes. The types of chemicals that have proven to be the greatest problems seem to be those that process the properties of lower water solubility, a higher octanol/water partition coefficient, longer as opposed to shorter chain lengths, environmental persistence and partitioning to solids.

The challenge before us is to get the data that will validate our laboratory acute and chronic toxicity data. This validation can be at least partially achieved if the opportunity is seized to look back in 20/20 hindsight into our many data bases that are available on existing chemicals.

4.4.5.4 Environmental Decison. QSAR discussed in this chapter will be supplemented by QSAR dealing with environmental fate. Figure 1 shows the parallel relationship between estimating environmental no-effect concentrations and exposure concentration. The difference between these two values is the safety margin that may be used in environmental decisions. Until some experience is gained in predicting effects and fate of chemicals by QSAR, regulatory decisions should not be based solely on QSAR predictions. It is recommended that existing regulations be evaluated by QSAR predictions to gain experience in the process.

GROUP V

TABLE OF CONTENTS

Page

GROUP V: STRUCTURE-ACTIVITY APPLICATION IN REGULATION

Summary 1

Introduction 3

Historical and Current Uses of Structure-Activity Relationships 9

Uses and Limitations of QSAR in Environmental Evaluations 20

Recommendations 23

References 25

GROUP V: STRUCTURE-ACTIVITY APPLICATION IN REGULATION

Summary

With many thousands of compounds about which little is known as to their environmental impact, there is need of rapid cost effective and reliable methods of evaluation to serve as a basis of regulation.

There is clear language in TSCA and its legislative history which authorizes EPA to delineate categories of chemicals which are similar in molecular structure, in physical, chemical, or biological properties, in use, or in mode of entrance into the human body, or the environment.

Review of SAR, as a tool in regulation of chemicals, reveals that it is a technique of proven predictive potential in some areas such as physical properties. It is currently being used by EPA as a tool to identify chemicals of possible concern, and exclude those of little concern, particularly where other experimental data are lacking.

The QSAR approach, while promising, cannot in all cases be relied upon as the sole basis for regulation. As a rule, experimental data are the preferred basis for regulatory decision making; however, in their absence, SAR can often provide some needed predictive capabilities. In the absence of health and environmental data, OTS already utilizes SAR in its premanufacture notice review program.

In recognition of the potential value of SAR in evaluating possible hazard of chemicals, it is recommended to EPA and others, to compile the available information into a computer data base to ensure the fullest extent of its utilization.

To assure continued development of SAR as a general and reliable technique of evaluating chemicals for regulatory purposes, it is recommended to the Agency that it should encourage and support increased research and development.

Introduction

There are some chemicals and intermediates in commerce manufactured in volumes ranging from pounds to billions of pounds per year. The demand for the chemicals is in a state of dynamic flux that leads to reduction in volume or complete disappearance from the market for some, and the introduction of new, chemicals each year. Generally, the introduction of new chemicals is larger than the disappearance of old or out-dated chemicals, thus leading to a growing number of commerce.

The quantity and nature of these chemicals almost inevitably results in some human exposure and in environmental contamination (Eto 1974). Obviously it is necessary, if unacceptable environmental impacts of these chemicals are to be avoided, that sufficient information must be available concerning their environmental fate and biological effects to develop a better system of manufacture, handling, use, and disposal.

The rapid growth in the number of chemicals used by our society and the late development of the scientific knowledge regarding possible adverse environmental impacts, has left us with a lack of information about many of these chemicals. With the recognition of the lack of this knowledge has come the determination to develop a rational basis for manufacture and use to avoid unacceptable adverse consequences. This is demonstrated in the passage of the Toxic

Substances Control Act (94th Congress 1976) and a number of other acts providing for the regulation of chemicals in the environment.

The passage of laws regulating chemicals indicates the desire to enjoy the benefits that accrue from use of the chemicals while minimizing the unwanted effects. As a consequence, both laws and promulgated regulations under the laws attempt to employ the best available information as a basis of judgment of safe manufacture and use of the chemicals. Thus, information as to the substance's chemistry, biological effects, method of transport, use, and disposal are always brought into regulatory consideration. However, the data base, particularly that of the chemistry in the environment and biological effects, often contain serious gaps. In part this is due to the methods that must be used in acquiring the information. For example, to evaluate the biological effects of a single chemical on a number of organisms at several different exposure levels, would entail an effort of perhaps as much as five work years and several hundred thousand dollars. Similarly, the determination of the environmental fate and behavior would involve nearly as much time and effort. Thus it is, that for many of the 45,000 chemicals in commerce, there is little of this type of information that has been developed.

Over the years, a number of tests have been developed to assess the biological effects and environmental fate and behavior of chemicals (National Academy of Sciences 1975; Kenega 1978; Metcalf 1973). Most of the biological tests have relied on exposure of an organism to several different dosages of chemicals administered by one or more routes, and examining the effect at varying periods of time. Simiarly, the environmental transport of a chemical often has been detected by monitoring efforts that yield an answer only sometime after the escape or release of the chemials.

The types of organisms used to evaluate biological effects range from bacteria and unicellular organisms to insects, fungi, plants, rodents, on up through primates. While some of the tests are relatively rapid, others require months or even years for completion. Obviously the time and cost in manpower and money make it prohibitive to do a thorough evaluation of all the chemicals currently in use, or that may be introduced in the future.

It is important to note that not all chemicals pose the same magnitude of problems. In fact many, for example polymers, are relatively inert, causing little or very small biological effects. Likewise, their environmental impact can be readily minimized. Therefore, it would be highly desirable to have a scheme of categorization and prioritization that would enable us to distinguish between the chemicals posing few, if any problems, from those that are potentially troublesome. With such a scheme, the attention and resources could then be focused on the problem chemicals and the development of a rational procedure of manufacture, use, and disposal. It is on this basis that the concept of the hierarchy of testing came into being. Even so, if in the hierarchy of testing, long-term biological assays or monitoring are required, it still becomes difficult and time consuming to identify the more troublesome of the chemicals for further attention. For that reason, a more rapid method of identifying those chemicals is required.

In the nineteenth century, scientists investigating the biological activity of chemicals developed a concept that the active chemicals reacted with receptor sites in the living organism, and that in order to do so must have a spatial configuration and physical properties that accommodate to the receptor site. This was a "lock and key," Ehrlich. Subsequently, it was observed that with narcotic agents certain properties, notably water solubility, vapor pressure and partition coefficient, were associated with that activity. Based on these early observations, the chemists and biologists postulated that there was a necessary and sufficient relationship between the chemical and physical properties of the compound and the biological effects produced (Sexton 1963; Albert 1973). A great deal of research effort has gone into attempting to elucidate an understanding of this relationship and generalize it to fit all instances where compounds exhibit biological activities. It was hoped that from this, the biological activity of a new structure could be predicted, thus facilitating both the synthesis of new compounds of utility and the development of appropriate methods of managing existing compounds.

Efforts have increasingly turned to devising quantitative structure-activity relationships by which one could predict the particular biological response that would be produced by a chemical. Similarly, this information would also be used in assessing probable environmental fate and behavior. The so-called quantitative structure-activity relationship (QSAR) in fact encompasses a broader concept thatn structure, inasmuch as it is concerned with both chemical and physical properties (Albert 1975).

The QSAR approach involving both chemical and physical properties has become increasingly sophisticated thanks to the efforts of a few dedicated investigators. From its early beginnings where attempts were made to relate simple elemental structure, molecular weight, vapor pressure, water solubility, and partition coefficient to biological activity, it has been elaborated to utilize the much more refined and sophisticated chemical concepts of kinetics and quantum mechanics. While the most successful applications have utilized the linear free energy approach such as that based on partition coefficient coupled with the Hammett constant, other approaches utilizing molecular orbital calculations, cluster analyses and pattern recognition have all served to increase the power of QSAR (Topliss 1978; Helmer 1968).

In considering the priority problem of developing rational guidelines and regulations for the safe manufacturing and use of chemicals, it would appear that few techniques of evaluation offer more potential value than QSAR.

While newer techniques and approaches to evaluation may develop in the future, at the moment it would appear that this method offers sufficient promise, that those applicable QSAR techniques should be utilized and as the field advances further, the new development adapted as quickly as possible to the regulatory procedures. Indeed, it is felt in many quarters that at the moment, QSAR offers perhaps the major, if not the only means of quickly evaluating the environmental impact and possible biological effects of the host of chemicals employed by a technological society.

At the moment, QSAR's greatest value is in making these evaluations in the absence of empirical data. Also in the case of compounds possessing significant potential biological activity, appropriate biological and environmental investigations would be required to provide the refined answers needed. At the moment, QSAR is not sufficiently developed to be more than an effective preliminary screening tool where experimental data are lacking.

It is the purpose of this section then to address the application and limitations of QSAR in relation to the regualtion of chemicals.

Historical and Current Uses of Structure-Activity Relationships (SAR)

A. Technical

The use of structure-activity relationships (QSAR) as a guide to the design of biologically active chemicals is not new. In the late 19th century, observations were made concerning the lipophilicity of pharmaceutical molecules and their narcotic activity (Sexton, 1963; Albert, 1973). Structure has now been related to toxicity, and the physical and chemical properties of compounds. These latter properties are importantly related to hazard or risk assessment since they relate to the value required to predict probable exposure, fate, and toxicity of chemials in the environment. Prediction of physical properties is perhaps the most successful use of SAR, since, by means of correlative values such as molecular additivity, pattern recognition, etc, from a number of chemicals, basic properties such as water solubility, vapor pressure, octanol-water partition coefficient can be predicted.

Based upon water solubility, molecular weight, vapor pressure and Henry's law, the partitioning coefficients between two media such as air:water, air:soil, soil:water, water:fat, etc., can be predicted under equilibrium conditions with accuracy needed to determine the major media (or compartment) into which the chemical will distribute when released in the natural environment (Tulp 1978).

To date, our knowledge and ability for predicting biological activity and environmental fate is based on chemical structure, but is fragmentary. The rate of hydrolysis has been shown to be directly correlated with cholinesterase inhibition and has proven useful in predicting the activity of organophosphate insecticides (Eto, 1974). This could indicate that good CHE inhibitors would likely undergo rapid hydrolysis in the environment.

Use of Hammet sigma constant as a prediction of electron withdrawing or donating properties of a substitute, has also been used in designing new biological active compounds. In pesticide chemistry, it has been used for predicting rates of dehydrochlorination of DDT analogs and therefore could play a very significant role in predicting environmental fate of a given molecule.

Photolytic degradation half life of a chemical is predictable, based on the functional groups and absorption spectrum of the compound itself. Research in this area is in progress and hopefully in time, will become a very useful tool.

Prediction of metabolic pathways of many newer chemicals, based on the knowledge of metabolism of known compounds, is not too far out of reach. Pathways of degradation in general are well known, and if a new chemical has substitutes or molecular regions which can be attacked, then it should be easy to predict its metabolic fate.

The toxicity of chemicals based on SAR is much more complex than predicting physical or chemical properties. Toxicity is a combination of the correct physical and chemical properties allowing the chemical to penetrate various tissues of organisms, be stable enough to reach the sensitive site, and exert specific toxic action, whatever it may be. SAR is useful in indicating the probable toxicity of large classes of chemicals such as organophosphate or carbamate insecticides, quite good for generic groups of chemicals such as substituted phenyl carbamates, and is of more limited values for specific chemicals.

B. Regulatory

1. Statutory and Other Authorities Supporting Use of Structure-Activity Relationships, TSCA:

The Toxic Substances Control Act (TSCA) authorizes the administrator of EPA to take any action under the act with respect to a "Category of Chemical Substances or Mixtures" (15 U.S.C. 2625$^{(e)}$). Categories of substances or mixtures are defined as groups, the members of which are similar in molecular structure, in physical, chemical, or biologial properties, in use, or in the mode of entrance into the human body or the environment. Although the act does not specify procedures for establishing categories, it is clear that structure-activity relationships would facilitate the regulation of such groups of chemical substances or mixtures.

Discussions of chemical categorization can be found in the various House and Senate Committee reports accompanying TSCA. For example, the House Report indicates that testing requirements can be imposed by EPA on two whole categories of chemicals at a time. The categorization approach would permit EPA to regulate groups of chemical substances of mixtures without the necessity for the requisite finding of harm for each member chemical and the category (H.R. Rep. No. 1341, 94th Congress., 2nd Session 61, 1976).

Congress intent was therefore, to require EPA to establish regulations expeditiously for control of the vast volume of chemical substances and mixtures which have some into use and become a pervasive part of our environment. This is apparent from the frequent reference in TSCA to control of categories and substances. In the absence of specific information on every individual substance or mixture in a category, the use of structure-activity relationships can fulfill the congressional goals established in the Act.

The legislative history of TSCA cites with approval, a National Academy of Sciences (NAS) (4) Report which recommends that EPA adopt a matrix or generic approach in regulating chemicals in the environment (NAS 1975a). The report states philosophically that rational regulatory priorities can be established with the use of generic approaches since logical relationships among the physical and chemical properties of, and groups of substances can be inferred from the known physical and chemical properties of one or more members of the group of substances. A second NAS study report, issued simultaneously, recommended that guidelines be established for classifying chemicals into categories so as to facilitate the study of their environment's effects (NAS 1975b). This report recognizes that to study every chemical to the same extent would represent an unjustifiable expenditure of effort that would not contribute significantly to the public health.

2) Selected Current Applications of SAR by OTS:

a) PMN Review Process - Section five of TSCA requires anyone who intends to manufacture a "new" chemical (i.e., are not appearing on the TSCA Section 8(b) Inventory) for a commercial purpose must notify (premature notification or PMN) EPA at least 90 days in advance of the intended date of manufacture and must also provide certain types of information to the EPA. TSCA does not require the manufacturer to conduct health and safety studies on the new chemical, although the manufacturer must provide to EPA all such data in their posession. Because the amount of physical chemical health and safety data provided in each PMN varies greatly from nothing more than chemical structures at one extreme to fairly detailed submissions of test data at the other extreme, EPA has chosen to employ SAR techniques in an attempt to predict the potential environmental fate and effects of new chemicals. A description of the OTS PMN review process follows:

The Office of Toxic Substances convenes a meeting of the Structure-Activity Team (SAT) to consider the fate and effects of each PMN chemical received since the last meeting. The SAT is composed of OTS' senior scientists having expertise in the disciplines of chemistry, environmental fate assessment, medicine, pharmacology, toxicology, and ecotoxicology. Using a comibination of professional expertise, the information provided in the PMN and information which is available in standard references (Merck Index, Goodman and Gilman, etc.), the SAT provides its best estimate of the fate and effect of the chemical. This process can best be described in SAR terms as one utilizing the technique of pattern recognition. When concern is expressed about the chemical, the SAT also provides guidance for the selection of structural analogs of the PMN chemical for comparison. The deliberations of the SAT are then summarized in an "SAT Report" for each PMN chemical and, where indicated, attempts are made to identify appropriate analogs for literature searching purposes. The tool used most commonly for the latter purpose is the CIS Structure and Nomenclature Search System (SANSS), which permits analysis of its files by substructure searching techniques. In this context, an "appropriate analog" is one which includes both the essential substructural features of the PMN chemical and offers the promise of a productive literature search. Analog selection is perhaps the most critical and the most difficult step in the PMN assessment process and is at present more an art than a science.

Following selection of analogs, a literature search is performed, the retrieval abstracts are screened for hard copy acquisition, and the literature is collected. Using the result of the literature search and the guidance provided by the SAT, a staff scientist prepares an Initial Screen Report which examines in as much detail as possible the environmental fate, the human health effect, and the environmental effects of the PMN chemical. Much of this work involves assessment by analogy to existing data on similar chemical structures and often tends to be qualitative in nature. The process to this point requires approximately 20 of the 90 days in the review period.

The Initial Screen Report is then combined with reports on the chemistry (primarily production and process) and release exposure of the chemical to yield an overall risk estimation. At this point a decision is made to, (1) drop the chemical from further consideration; (2) select it for a possible Significant New Use Rule (SNUR), under section 5 (a) (2); or (3) put it through "Detailed Review" in order to further assess the chemical and the potential need for regualtion. If the last choice is selected, the areas of greatest concern are identified and subjected to an in-depth assessment conducted by appropriately trained scientsits in OTS. The "Detailed Review" scientists closely examine the pertinent portions of the Initial Screen Report, and using validation procedures for both the data, and to the extent possible, the SAR arguments, attempt to confirm and expand on the tentative conclusion offered in the Initial Screen Report. The "Detailed Review" scientists may also be asked to evaluate potential substitutes for the PMN chemical in terms of the relative concern attached to each. Following detailed examination of the available data and careful consideration of any SAR based arguments, the "Detailed Review" scientists prepare an assessment which documents the scientific basis for their concerns. This document then serves as the technical basis for the regulatory decision that must be made at this point as well as the scientific support of any subsequently promulgated regulations.

b) Illustrative Regulatory Action: TSCA, Section 5 (e)

Where the screening and detailed review of a new chemical discussed above reveal significant health or environmental concerns, TSCA provides EPA a number of regulatory options. In instances where the Agency has insufficient data and has relied on an SAR approach, the available regulatory controls will of necessity derive their scientific basis from the structure-activity analysis.

For example, EPA has authority under TSCA, Section 5 (e) to regulate the introduction of new chemicals of concern pending the development of information sufficient to assess the chemical's health and/or environmental effects. The Agency has issued two proposed administrative orders under that provision and is currently negotiating others in the form of "consent orders". The structure-activity analyses conducted during the notice review process, in conjunction with the relevant exposure projections and numerous other regulatory considerations, have constituted the basis for these orders and negotiated controls.

The initial proposed Section 5 (e) order, issued on April 23, 1980, against a group of phthalate polymers, utilized structure-activity analysis in conjunction with a National Cancer Institute (NCI) study on a close structural analog. The new substances possessed strong structural and chemical similarities to the NCI test chemical, diethylhexylphthalate (DEHP). The NCI test data strongly indicated that DEHP was a liver carcinogen in mice and rats. The structural similarity of the PMN phthalates to DEHP, and the expectation that the PMN chemicals and DEHP share similar or common metabolites, indicated that the PMN chemicals in question might also may be carinogenic. Other aspects of the proposed Section 5 (e) order which relied on SAR considerations included, (1) the high potential for bioconcentration of the PMN chemicals as indicated by log P calculations and structural similarity to known bioaccumulative phthates; (2) the structural similarity of the PMN chemicals to longer chain alkylphthalates biodegrade slowly under anaerobic conditions and strongly adsorb to soils and sediments indicated that the PMN chemicals may persist in the environment for a significant period, thus creating the possibility of long-term exposure to aquatic organisms.

A second proposed order under section 5 (e) addressing human health concerns was issued on September 5, 1980. Substructural analysis of a new chemical whose specific structure was claimed confidential, revealed the existence of a close structural analog with well-documented sensitization and irritation effects in both humans and animals. There was no data available to EPA indicating that the minor structural differences between the two chemicals would mitigate the concerns raised by the close analog. The sensitization concerns coupled with widespread human exposure and other factors entered into the "unreasonable risk" determination.

Although the initial two section 8(e) orders resulted in withdrawal of the proposals to manufacture, a number of companies are expressing a desire for negotiated controls. Here again the use of SAR during Agency review has played a role in the case developments, as indeed it must in the assessment of new chemicals.

c) Other Uses in OTS

The use of SAR techniques have been found useful in several other activities relating to TSCA. The Interagency Testing Committee has utilized SAR in its procedures for the identification and selection of chemicals which should be considered for testing under TSCA section 4 as well as the selection of the specific effects or categories of effects that require testing.

SAR has also been used as a basis for the selection of potential candidates for the first stage of risk assessment under section 6, namely preparation of a "Chemical Hazard Information Profile" (CHIP). In a similar fashion, SAR also holds promise as an aspect in the "scoring" of chemicals in distinguishing those which should be examined in more detail from those which may not be candidates for regulatory action. For example, log P calculations can be used to distinguish those chemicals having a potential for bioaccumulation from those which, on a basis of their log P values appear less likely to be bioaccumulative. SAR is useful in selecting and ranking chemicals subject to possible "significant new use" rules under TSCA section 5 (a) (2).

There are, no doubt, additional instances throughout EPA where SAR has been or could be a useful device in the assessment and ranking of chemicals. These examples are given merely to illustrate, in part, its current practical applications.

Uses and Limitations of QSAR in Environmenal Evaluations

At the present time, the major limitation of SAR is the inadequacy of the data necessary to fully utilize it. While volumes of data have been generated over many years, much of it, particularly that generated by industry, is not available at all or at best is not available in a readily and easily accessible form. Few computerized data bases are currently in existence, all of them incomplete - I SHOW, PROPHET, AND HEEDA for example. Much more input into these and other systems is necessary to fully utilize SAR as a valuable tool.

A second limitation is reliability of predictions based on mathematic or structural models. In developing such models, validation is needed to correlate the numerical values generated with values determined in actual experiments. While many of these correlations have been developed, such as that of log P with bioconcentration, and calculated hydrolyis rate constants with laboratory-determined rates, few other models have been validated.

The validation described above does not take into account the errors associated with the prediction, nor its accuracy and applicability in field situations. The applicability of SAR use is only as good as the range of standard errors associated with the descriptor being employed. Thus, the needs of the user - manufacturer for screening or regulator for policy setting - must be fully understood before applying SAR in decision making.

285

The degree of reliability can vary with the use. A lower degree of reliability may be acceptable for decision tree analyses, while a much greater degree of certainty must be used in regulation development. A great degree of certainty must be shown in order to ban a chemical in order to sustain both scientific and judicial review. On the other hand, to prioritize chemicals for evaluation or to screen candidates, new products may not require such stringent criteria. In many cases of prioritization, one has no choice but to use SAR as the only decision making tool.

To date, the use of SAR has been limited to new product development and to a lesser extent, the PMN process. Other modeling techniques have been utilized in the regulatory process andhave withstood the adjudicatory process.

As stated earlier, SAR has historically been used in new product development. The pharmaceutical industry has used chemical structure in predicting the efficiency of drugs; as has also the pesticide industry (11). It has been shown that SAR can predict some of the properties which a compound will exhibit. This can include hydrophilicity, conductivity, and bioconcentration potential.

In the area of environmental effects, SAR is showing promise. For example, from the physical properties of water solubility or log P, an estimate of bioconcentration potential can be derived. From log P, the acute lethality value (LC50) can presently be estimated for at least six classes of chemicals. For those same classes of chemicals, their acute lethality can also be estimated directly from molecular structure using Kier-Hall connectivety indices.

In the area of environmental fate and transport, several examples can be cited for advantageous use of SAR. From molecular structure, air and water oxidation rates can be derived, then used for estimating the persistence of a compound. Soil or sediment adsorption can be estimated from the Kow; arrived at from the molecular structure by use of additivity rules. Likewise, the surface of a molecule can be used to derive a compound's water solubility, indicative of its media partitioning.

SAR is expected to be utilized in establishing agency priority in addressing the sorting, evaluation, and control of chemicals. In priority setting, order of magnitude values are often adequate. SAR will also be utilized in evaluating a given industrial facility (air/water/residual) for potential human health or environmental harm. As example, NPDES permits may be based on predicted LC50 values. Proposed regulatory decisions for hazardous wastes, air emissions, water discharges, etc. could, where adequate data does not exist, utilize SAR predictions of toxicity, persistence, bioaccumulation, etc.

Recommendations

Review of SAR as a tool in regulation of chemicals reveals that it is a technique of proven predictive potential in some areas, such as physical properties. It is currently being used by EPA particularly in cases where available as a tool to identify chemicals of possible concern and to exclude those of no concern. It should not be used in regulatory areas where science does not support their use.

In recognition of the value of SAR, despite its limitations, it is recommended to the agency that: (1) they compile the available physical, chemical, and biological information on a number of chemicals, including benchmark chemicals, into a computerized data base; (2) they evaluate available data for validity to establish needed correlation on a base set of chemicals for predictive purposes where needed. Further, to attempt to fill data gaps in this information through reliable laboratory and field methods; (3) they recognize that SAR is only one of the set of tools for evaluating chemicals, the real endpoint being determination of the margin of safety between concentrations toxic to organisms and concentrations occurring in the environment from all sources; (4) they recognize that SAR is a valuable tool and its use should be encouraged in tiered testing and for prioritization of chemicals in regulation, and for assessing new chemicals where data are often lacking; (5) they recognize the limitations of precision and accuracy of the methods used and the reliability of the predictions made regarding biological effects and environmental impacts when using current SAR's as an internal basis for regulation; (6) they should further the development of SAR through increased encouragement and support of research for applications in regulation of toxic and hazardous chemicals.

REFERENCES

Albert, A. 1971. Relations Between Molecular Structure and Biological Activity: Stages in the Evolution of Current Concepts. In: Annual Review of Pharmacology, ed. H.W. Elliott. pp. 13-36. Annual Reviews, Inc. Palo Alto, California.

Albert, A. 1973. Selective Toxicity, The Physico-Chemical Basis of Therapy. Chapman and Hall Ltd. London.

Eto, M. 1974. Organophosphorus Pesticides: Organic and Biological Chemistry. CRC Press, Inc. Cleveland, Ohio.

Helmer, F., Kiehs, K. and Hansch, C. 1968. The Linear Free-Energy Relationship Between Partition Coefficients and the Binding and Conformational Perturbation of Macromolecules by Small Organic Compounds. Biochm. 7:2858-2863.

Kenega, E.E. 1978. Test Organisms and Methods Useful for Early Assessment of Acute Toxicity of Chemicals. Environ. Sci. Tech. 12:1322-1329.

Metcalf, R.L., Kapoor, I.P.; Hirwe, A.S.; Coats, J.R.; and Khalsa, M.S. 1973. Structure Activity Correlations of Biodegradability of DDT analogs. J. Agric. Food Chem. 21:2.

National Academy of Sciences. 1975a. Decision Making for Regulating Chemicals in the Environment. Washington, D.C.

290

National Academy of Sciences. 1975b. Principles for Evaluating Chemicals in the Environment. Washington, D.C.

Public Law 94-469. October 11, 1975. Toxic Substances Control Act. 94th Congress.

Sexton, W.A. 1963. Chemical Constitution and Biological Activity. D. Van Nostrand Company, Inc. Princeton, New Jersey.

Topliss, J.G. and Fukunaga, James Y. 1978. Chapter 30. Quantitative Structure-Activity Relationships in Drug Design. Annual Reports in Medicinal Chemistry. 13:292-303.

Tulp, M.TH.M. and Hutzinger, O. 1978. Some thoughts on Aqueous Solubilities and Partition Coefficients in PCB, and the Mathematical Correlation Between Bioaccumulation and Physico-Chemical Properties. Chemosphere 10:849-860.

WORKSHOP ON STRUCTURE ACTIVITY CONCEPTS IN ENVIRONMENTAL SCIENCES

February 1-4, 1981

EL TROPICANO HOTEL
SAN ANTONIO, TEXAS

PARTICIPATING OFFICES

Office of Pesticides and Toxic Substances

Office of Water and Waste Management

Office of Planning and Management

WORKSHOP CHAIRMEN

Rizwanul Haque
Office of Environmental Processes
and Effects Research

Gilman Veith
Environmental Research Laboratories
Duluth, Minnesota

WORKSHOP COORDINATION

The MITRE Corporation
McLean, Virginia

SPONSORED BY

TOXICS AND PESTICIDES DIVISION

William S. Murray, Ph.D., Director
Office of Environmental Processes
and Effects Research

Office of Research and Development

U. S. Environmental Protection Agency

PARTICIPANTS
EPA WORKSHOP ON STRUCTURE ACTIVITY CONCEPTS IN ENVIRONMENTAL SCIENCES

Dr. Etcyl H. Blair
Vice President
Director, Health and Environmental
  Sciences
The Dow Chemical Co.
Midland, Michigan 48460
(517) 636-5509

Dr. Corwin Hansch
Pomona College Medicinal Chemistry Project
Seaver Chemistry Laboratory
Claremont, California 91711
(714) 621-8000 ext 2225

Dr. Rizwanul Haque
Office of Environmental Processes
  and Effects Research
Toxics and Pesticides Division
U.S. Environmental Protection Agency
401 M Street S.W.
Washington, D.C. 20460
(202) 426-0810

Dr. William Murray
Toxics and Pesticides Division
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, D.C. 20460
(202) 426-2760

PARTICIPANTS
EPA WORKSHOP ON STRUCTURE ACTIVITY CONCEPTS IN ENVIRONMENTAL SCIENCES

GROUP I - AVAILABLE METHODS FOR STRUCTURE ACTIVITY - REQUIREMENT AND LIMITATIONS

Dr. Stephen Lozano — Co-Chairperson
Environmental Research Lab
6201 Congdon Blvd.
Duluth, MN 55804
(218) 727-6697 ext. 568
FTS: 783-9568

Dr. Robert Lipnick — Co-Chairperson
Environmental Protection Agency
Office of Toxic Substances
401 M Street, S.W.
Washington, D.C. 20460
(202) 472-3340
FTS: 755-4880

Dr. William Dunn
University of Illinois Medical Center
833 So. Wood
Chicago, Ill.
(312) 996-5232

Mr. Kurt Enslein
Health Designs, Inc.
Suite 1428
183 Main Street, East
Rochester, N.Y. 14604
(716) 546-1464

Dr. Martin Hamilton
Department of Math
2-214 Wilson Hall
Montana State University
Bozeman, Montana 59717
(406) 994-3601

Dr. Peter Jurs
Penn State University
152 Davey Lab.
University Park, PA 16802
(814) 865-3739

GROUP I (cont)

Dr. Yvonne Martin
Medicinal Chemistry Dept.
Abbott Laboratories
Department 466
N. Chicago, Ill. 60064
(312) 937-5362

Dr. Rudy Potenzone
Environmental Protection Agency
MIDS, PM-218
401 M Street, S.W.
Washington, D.C. 20460
(202) 245-3027
FTS: 245-3027

Dr. William Purcell
P.O. Box 41151
University of Tennessee
Memphis, Tenn. 38104
(901) 528-6050

GROUP II

EVALUATION OF MOLECULAR DESCRIPTORS

Dr. Albert J. Leo Co-Chairperson
Chemistry Department
Pomona College
Claremont, California 91711
(914) 621-8000 ext 2225

Dr. Bruce W. Gay, Jr. Co-Chairperson
Environmental Research Laboratory
U.S. Environmental Protection Agency
Mail Drop 84
Research Triangle Park, N.C. 27711
(919) 541-2830
FTS: 629-2830

Dr. Marvin Charton
Pratt Institute
Chemistry Department
Brooklyn, N.Y. 11205
(212) 636-3763

Dr. Richard Cramer
Smith, Kline Corporation
Box 7929
Philadelphia, PA 19101
(215) 854-5648

Dr. Lowell H. Hall
Eastern Nazarene College
23 E. Elm Avenue
Quincy, MA 02170
(617) 773-6350

Dr. Jack Hine
Chemistry Department
Ohio State University
180 W 18th Ave.
Columbus, OH 43210
(614) 422-7676

Dr. Gilda Loew
SRI/International
333 Ravenswood Avenue
Menlo Park, CA 94025
(415) 859-3105

Dr. Robert Pearlman
College of Pharmacy
University of Texas
Austin, Texas
(512) 471-3383

GROUP II (cont)

Dr. David Weininger
Environmental Research Laboratory
Environmental Protection Agency
6201 Congdon Blvd.
Duluth, MN 55811
(218) 727-6692 ext 568
FTS: 783-9568

Dr. Samuel Yalkowsky
The Upjohn Co.
Pharmaceutical Research Unit
Kalmazoo, MI 49001
(616) 323-5702

GROUP III

ENVIRONMENTAL FATE PREDICTION

Dr. Theodore Mill SRI/International PS-273 333 Ravenswood Avenue Menlo Park, California 94025 (415) 859-3605	Co-Chairperson
Dr. George Baughman Environmental Research Laboratory Environmental Protection Agency College Station Road Athens, GA 30613 (404) 546-3145 FTS: 250-3145	Co-Chairperson

---

Dr. Judith C. Harris
A.D. Little, Inc.
15-311 Acorn Park
Cambridge, MA 02140
(617) 864-5770 x 3092

Dr. Dale Hendry
SRI/International
Physical Organic Chemistry Department
333 Ravenswood Drive
Menlo Park, CA 94025
(415) 859-2015

Dr. Don MacKay
Department of Chemical Engineering
University of Toronto
Toronto, ONT
CANADA
(416) 978-4019

Dr. Authur E. Martell
Texas A&M University
Department of Chemistry
College Station, TX 77843
(713) 845-5055

Dr. I. Suffet
Environmental Studies Institute
Drexel University
32nd and Chestnut Streets
Philadelphia, PA 19104
(215) 895-2270

GROUP III (cont)

Dr. Perry McCarty
Stanford University
Department of Civil Engineering
Stanford, CA 94305
(415) 497-3921

Dr. David H. Rosenblatt
U.S. Medical Bioengineering
Research and Development Laboratory
Building 568
Ft. Detrick
Frederick, MD 21701
(301) 663-2014
FTS: 935-2014

Dr. R. Srinivasan
IBM Research Center
P.O. Box 218
Watson Research Center
Yorktown Heights, N.Y. 10598
(914) 945-2084

Dr. A. M. Stern
Environmental Protection Agency
Office of Toxic Substances TS-798
401 M Street, S.W.
Washington, D.C. 20460
(202) 426-9145
FTS: 8-426-9145

GROUP IV

ENVIRONMENTAL EFFECTS PREDICTION

Dr. Dean Branson — Co-chairperson
Health and Environmental Sciences Research
Dow Chemical Co.
5002 Highridge Ct.
Midland, MI 48640
(517) 636-2730

Dr. Gilman Veith — Co-chairperson
Environmental Research Laboratory
6201 Congdon Blvd.
Duluth, Minn. 55804
(218) 727-6692 ext. 534
FTS: 783-9534

Dr. Rolf Hartung
Department Environmental and
Industrial Health
University of Michigan
Ann Arbor, MI 48109
(313) 764-5430

Dr. Lemont B. Kier
School of Pharmacy
Medical College of Virginia
Box 581
Richmond, VA 23298
(804) 786-0650

Dr. Richard Kimerle
Monsanto Industrial Chemical Co.
800 N Lindbergh Blvd.
St. Louis, Missouri 63166
(314) 694-3286

Dr. Alan W. Maki
Proctor and Gamble Co.
Environmental Safety Dept.
Ivorydale Technical Center
Cincinnati, OH 45224
(513) 763-6339

Dr. Craig MacFarlane
Environmental Research Laboratory
Environmental Protection Agency
200 S.W. 35th Street
Crovallis, Oregon 97330
(503) 757-4670
FTS: 420-4670

GROUP IV (cont)

Dr. Foster Mayer*
Columbia National Fisheries Research Laboratory
Department of Energy
Fish & Wildlife Service
Rt. 1
Columbia, Missouri 65201
(314) 442-2271

Dr. James Lake
Environmental Protection Agency
Environmental Research Laboratory
S. Ferry Road
Naragansett, RI 02882
401-789-1071
FTS: 838-4843

Dr. Terry Wayne Schultz
Oak Ridge National Laboratory
Biology Division
P.O. Box Y
Oak Ridge, Tenn. 37830
(615) 574-1235
FTS: 624-1235

Dr. Robert V. Thurston
Fisheries Bioassay Lab.
Montana State University
Bozeman, Montana 58717
(406) 994-3371

Dr. Valdo Zitko
Fisheries Environmental Sciences
St. Andrews, New Brunswick
EOG2XO Canada
(506) 529-8854

* Unable to attend - participated via conference telelphone hookup.

GROUP V

STRUCTURE ACTIVITY APPLICATION IN REGULATION

Dr. Virgil H. Freed Co-Chairperson
Oregon State University
Corvallis, OR 97330
(503) 754-3791

Ms. Laura Bennett Co-Chairperson
Environmental Protection Agency
Office of Environmental Processes
and Effects Research
RD-682
401 M Street, S.W.
Washington, D.C. 20460
(202) 426-0811
FTS: 426-0810

Mr. Charles Auer
Environmental Protection Agency
Office of Toxic Substances TS-792
401 M Street, S.W.
Washington, D.C. 20460
(202) 426-9819
FTS: 426-9819

Mr. Frank Biros
Environmental Protection Agency
Office of Hazardous Waste Enforcement
401 M Street, S.W.
Washington, D.C. 20460
(202) 426-8710

Ms. Rita Comoto (Bahner)
Assoc. Director for Health and Safety and Chemical Regulation
Chemical Manufacturer's Association
2501 M Street, N.W.
Washington, D.C. 20037
(202) 887-1282

Dr. Inder P. Kapoor
Manager, Plant Discovery
American Cyanamid Co.
P.O. Box 400
Princeton, N.J. 08540
(609) 799-0400

Dr. Eugene Kenaga
Health and Environmental Sciences
Dow Chemical Co.
Midland, MI 48640
(517) 636-1572

GROUP V (cont)

Mr. Glenn D. Pratt
Environmental Protection Agency
Region V Enforcement Division
230 South Dearborn Street
Chicago, Ill. 60643
(312) 886-6017
FTS: 886-6107

Ms. Cindy Work
Environmental Protection Agency
Office of Toxic Substances TS-794
401 M Street, S.W.
Washington, D.C. 20460
(202) 426-2601
FTS: 426-2601